

Klasifikasi Bahasa Isyarat Amerika menggunakan Convolutional Neural Network

Felicia Devina Siswanto^{a1}, Caecilia Citra Lestari^{a2}, Evan Tanuwijaya^{a3}

^aUniversitas Ciputra

CitraLand CBD Boulevard, Made, Kec. Sambikerep, Kota SBY, Jawa Timur 60219

¹fsiswanto@student.ciputra.ac.id

²caecilia.citra@ciputra.ac.id

³evan.tanuwijaya@ciputra.ac.id

Abstrak

Bahasa Isyarat adalah bahasa untuk orang - orang yang memiliki kesulitan mendengar maupun bicara. Tetapi bahasa isyarat bukanlah bahasa yang banyak digemari oleh masyarakat, sehingga orang yang memiliki disabilitas tersebut akan semakin kesulitan. Pada jurnal ini akan menjelaskan mengenai klasifikasi bahasa isyarat Amerika dengan menggunakan Convolutional Neural Network (CNN). Pada penelitian ini akan dilakukan beberapa penelitian menggunakan parameter berbeda seperti pada preprocessing, penelitian akan dilakukan dengan melihat parameter horizontal flip. Selanjutnya penelitian juga dilakukan dengan melihat epoch. Penelitian ini dilakukan untuk memantau akurasi dan akurasi validasi. Model yang dibuat pada penelitian ini nilai akurasi yang lebih tinggi saat memprediksi huruf v, dan n. Hasil nilai akurasi dari penelitian ini adalah 82.1%.

Kata kunci: Bahasa Isyarat, CNN, klasifikasi

American Sign Language Classification using Convolutional Neural Network

Abstract

Sign Language is a language for people who have hearing and speech difficulties. But sign language is not a language that is favored by many people, so people with disabilities will find it increasingly difficult. This journal will explain the classification of American sign language using the Convolutional Neural Network (CNN). In this study, several studies will be carried out using different parameters such as in preprocessing, research will be carried out by looking at the horizontal flip parameter. Furthermore, research was also carried out by looking at the epoch. This study was conducted to monitor the accuracy and accuracy of the validation. The model made in this study has a higher accuracy value when predicting the letters v, and n. The result of the accuracy value of this study is 82.1%.

Keywords: Sign Language, CNN, Classification

I. PENDAHULUAN

Bahasa Isyarat adalah bahasa yang digunakan oleh orang - orang yang memiliki kesulitan dalam mendengar maupun berbicara. Tak seperti bahasa di dunia, bahasa isyarat sering kali tidak diminati oleh khalayak luas.

Di Indonesia terdapat banyak sekali tuna rungu, dan tuna wicara. Menurut data pada tahun 2007, terdapat 16.335 anak penyandang tunawicara dan 3.861 anak penyandang tunarungu dan 7.632 anak penyandang tunarungu, tunawicara, dan tunanetra [1]. Banyak dari mereka menggunakan bahasa isyarat untuk berkomunikasi dengan orang lain. Tetapi karena minat orang - orang terhadap bahasa isyarat masih tidak cukup tinggi sehingga banyak orang tidak dapat memahami bahasa mereka.

Tuna Rungu juga seringkali kesusahan dalam memahami acara seperti debat atau berita karena jarang penerjemah bahasa isyarat. Selain itu penerjemah bahasa isyarat di Indonesia tidak banyak, salah satu contohnya ada di Sumatera Barat yang hanya memiliki 8 penerjemah bahasa isyarat untuk 1250 tuna rungu [2].

Pendeteksi pose tangan telah banyak digunakan dari waktu ke waktu [3]. Salah satu contohnya pada jurnal [4] yang menggunakan Doppler Radar dan VGG lalu mendapatkan akurasi sebesar 87.5%. Pada paper [5] klasifikasi dilakukan menggunakan pre train model deep neural network dan mendapatkan hasil 99%. Penelitian lain juga dilakukan untuk melakukan klasifikasi menggunakan kamera *handphone*. Klasifikasi tersebut mendapatkan 92.8% sebagai hasil akurasinya [6]. Pada

jurnal [7] dilakukan penelitian *static sign language recognition using convolutional neural network*, dan mendapatkan hasil 97.5%.

Oleh karena itu, fokus dari jurnal ini adalah untuk memberikan gambaran mengenai pengimplementasian deep learning *Convolutional Neural Network (CNN)* pada *American Sign Language (ASL) Classifier*. Adapun organisasi penulisan dari paper ini dibagi menjadi beberapa bagian yaitu metodologi, hasil dan pembahasan, serta kesimpulan.

II. METODOLOGI

Pada bab ini memuat langkah - langkah peneliti dalam melakukan penelitian secara lengkap dan padat. Penelitian akan dimulai dari penjelasan dataset, pemrosesan data, dan arsitektur yang digunakan

A. Dataset

Dataset yang digunakan adalah dataset kaggle dari Akash [8] berupa gambar pose tangan untuk bahasa isyarat. Dataset ini berjumlah 78000 gambar, berukuran 1.03 gb dan masing - masing gambar berukuran 200 piksel x 200 piksel. Pada dataset ini terdapat 26 kelas yaitu semua huruf alfabet mulai dari 'A' hingga 'Z'. Pada dataset ini terdapat 3 kelas tambahan yang berjudul 'nothing', 'del', dan 'space' tetapi karena tidak diperlukan maka 3 kelas tersebut tidak digunakan pada penelitian ini.

B. Preprocessing

Preprocessing gambar dilakukan dengan menggunakan library Image Data Generator. Image Data Generator adalah sebuah kelas pada keras yang digunakan untuk pipeline gambar untuk pembelajaran yang lebih dalam. Selain itu ImageDataGenerator dapat digunakan untuk pra processing data dan juga augmentasi data [9].

Image Data Generator memiliki banyak parameter dan beberapa parameter yang digunakan pada paper ini adalah

- `samplewise_center = True`,
- `horizontal_flip = True / False`,
- `samplewise_std_normalization = True`,
- `validation_split = 0.2`

Pada dataset ini dilakukan percobaan untuk melakukan preprocessing dengan data yang di horizontal flip dan yang satunya tidak dilakukan horizontal flip. Penelitian ini dilakukan untuk selanjutnya setelah menerapkan parameter yang tertulis maka akan di *flow* pada gambar training dan validation. Untuk training parameter *flow* nya adalah

- `target_size = (64,64)`
- `batch_size= 64`
- `shuffle = True`
- `subset = 'training'`

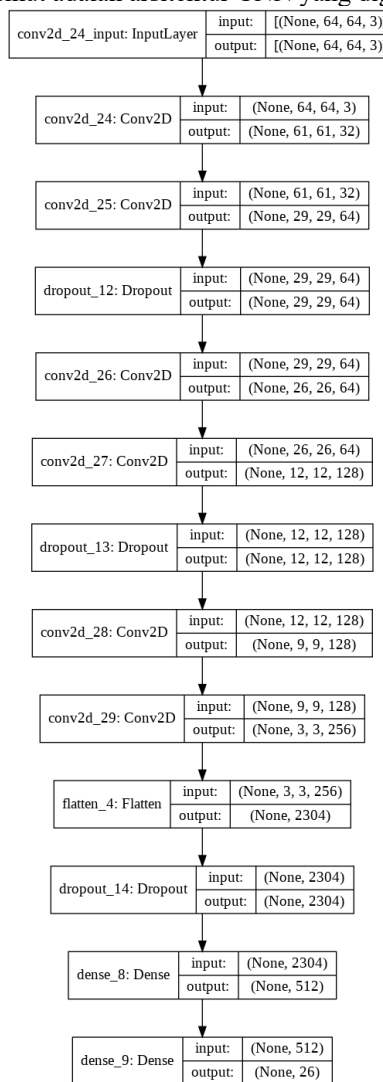
Sedangkan pada validation parameter flow nya adalah

- `target_size = (64,64)`
- `batch_size= 64`
- `shuffle = True`
- `subset = 'training'`

C. Preprocessing

Arsitektur yang digunakan pada jurnal ini adalah *Convolutional Neural Network (CNN)*. Beberapa referensi

digunakan dalam pembuatan arsitektur ini, yang pertama adalah dari jurnal berjudul "Sign Language Recognition System Using Deep Neural Network" [10], dimana digunakan 9 layer berupa Conv2D, pooling, dropout 2x lalu di flatten dan menambahkan 2 layer dense. Selanjutnya adalah dari jurnal berjudul "Using Deep Convolutional Networks for Gesture Recognition in American Sign Language" yang menggunakan 2 layer Conv2D, 1 layer Maxpool, dan 1 layer dropout sebanyak 3x lalu dilakukan flatten dan 1 dense, 1 dropout sebanyak 2x [11]. Berikut adalah arsitektur CNN yang digunakan



Gambar 1. Arsitektur model *Convolutional Neural Network (CNN)* yang digunakan dalam melakukan penelitian.

1) *Sequential*: Sequential adalah model pada keras yang digunakan untuk membuat model polos untuk layer. Model ini digunakan saat hanya memiliki 1 input dan 1 output. [12].

2) *Conv2D*: Conv2D adalah layer pada keras yang banyak digunakan pada arsitektur ini. Model dimulai dengan layer Conv2D dengan parameter filter 32, kernel_size 4, strides 1, activation relu dan input shape (64,64,3). Pada layer Conv2D berikut - berikutnya angka filter ditingkatkan menjadi 64, lalu 128, hingga 256.

Semua layer Conv2D menggunakan activation relu karena relu memiliki nilai terendah 0 sehingga semua angka dibawah 0 akan dianggap sebagai 0. Relu cocok digunakan untuk classifier ini untuk menghemat waktu pelatihan model [13].

3) *Dropout*: Dropout digunakan beberapa kali di arsitektur ini dan dilakukan drop yang pertama setelah 2 layer Conv2D dengan parameter rate 0.25, yang kedua juga menggunakan parameter rate 0.25, dan yang terakhir dilakukan dropout sebanyak 0.5. Dropout digunakan untuk menghindari overfitting. Dropout bekerja dengan cara menghilangkan beberapa unit selama proses pelatihan [14].

4) *Flatten*: Flatten digunakan 1x pada arsitektur ini yaitu sebelum dropout ketiga dan dense pertama. Flatten digunakan untuk mengubah array 2 dimensi menjadi 1 dimensi [15].

5) *Dense*: Dense akan dilakukan sebanyak 2x pada pembuatan arsitektur ini. Dense yang pertama menggunakan parameter 512, dan activation relu, sedangkan yang terakhir menggunakan parameter 26 sesuai dengan jumlah class atau folder pada dataset. Untuk yang layer dense kedua activation yang digunakan bukan relu lagi tetapi berupa softmax.

III. HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan hasil dari penelitian *American Sign Language Classifier* menggunakan beberapa indikator seperti pengaruh preprocessing data, dan epoch terhadap *accuracy*, *validation accuracy*, *loss* dan *validation loss*.

A. Preprocessing Data

Pada penelitian dilakukan percobaan pada preprocessing data. Pada saat menggunakan Image Library Generator parameter horizontal flip akan diteliti, apakah lebih baik jika digunakan ataukah tidak.

TABEL I
RATA - RATA AKURASI PADA PENGGUNAAN HORIZONTAL FLIP

Horizontal Flip	Rata - Rata Akurasi	Rata - Rata Akurasi Validation
True	0.95354	0,81718
False	0.97255	0.80322

Dari tabel tersebut dapat diketahui jika menggunakan horizontal flip akurasi yang didapat adalah 95.3% dan akurasi validation yang didapat adalah 81,7% sedangkan jika tidak menggunakan horizontal flip didapat akurasi sebesar 97.2% dan akurasi validation sebesar 80,3% tidak banyak perbedaan antara menggunakan horizontal flip maupun tidak. Tetapi akurasi lebih besar didapatkan saat tidak menggunakan horizontal flip.

B. Epoch

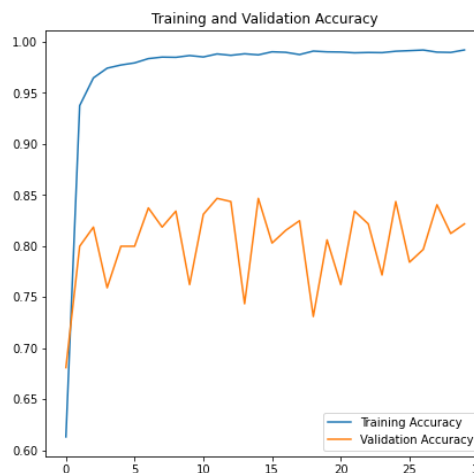
Penelitian juga dilakukan untuk mengetahui perbandingan nilai akurasi pada tiap perbedaan epoch.

TABEL II
PERBEDAAN AKURASI PADA TIAP EPOCH

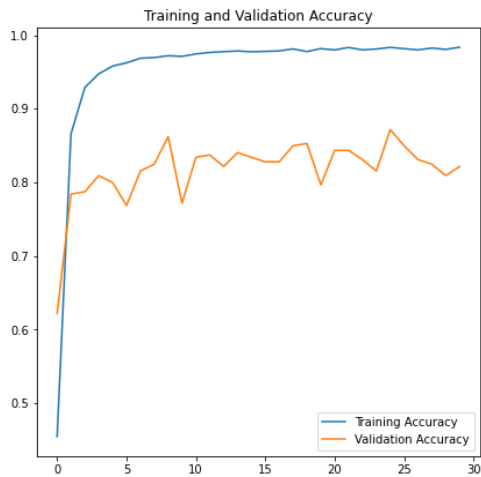
Epoch	Accuracy	Validation Accuracy
Menggunakan Horizontal Flip		
10	0.9715	0.7719
20	0.9820	0.7969
30	0.9839	0.8219
Tanpa Horizontal Flip		
10	0.9866	0.7625
20	0.9903	0.8062
30	0.9912	0.8219

Dari tabel dapat dilihat tidak banyak perbedaan pada tabel. Keduanya menampilkan akurasi rendah pada epoch ke-10 dan kenaikan pada epoch ke-20 dan pada epoch ke-30 keduanya mengalami kenaikan dan mendapatkan nilai yang sama yaitu 0.8219.

C. Accuracy dan Validation Accuracy



Gambar 2. Grafik training dan validation accuracy tanpa horizontal flip

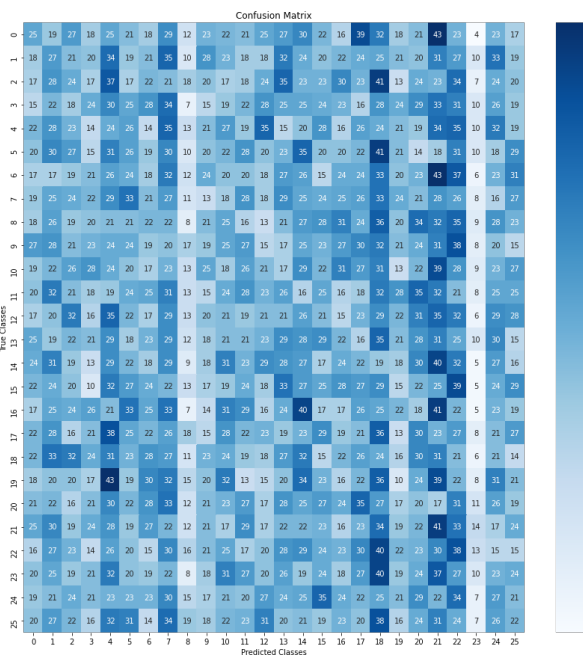


Gambar 3. Grafik training dan validation accuracy dengan horizontal flip

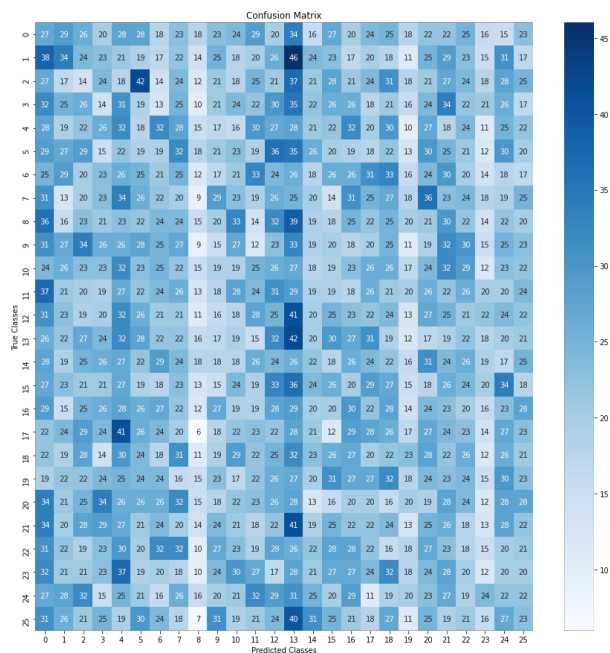
Kedua grafik tersebut, tidak menunjukkan banyak perbedaan accuracy pada penggunaan horizontal flip maupun tidak. Selain itu keduanya menampilkan grafik accuracy yang bagus pada training tetapi pada validation grafiknya tidak stabil.

D. Confusion Matrix

Judul menggunakan huruf Times New Roman reguler dengan ukuran 18pt. Nama penulis 11 pt. Afiliasi penulis 10 pt *Italic*. Alamat email menggunakan huruf reguler *Courier* dengan ukuran 9 pt.



Gambar 4. Hasil confusion matrix tanpa horizontal flip



Gambar 5. Hasil confusion matrix dengan horizontal flip

Jika dilihat dari hasil confusion matrix diatas dapat dilihat masih banyak kelas yang terprediksi salah saat validasi. Dapat dilihat dari warna biru yang masih tersebar luas.

E. F1 Score

	precision	recall	f1-score	support
a	0.05	0.04	0.04	600
b	0.04	0.04	0.04	600
c	0.04	0.04	0.04	600
d	0.05	0.04	0.04	600
e	0.03	0.04	0.04	600
f	0.04	0.04	0.04	600
g	0.03	0.03	0.03	600
h	0.04	0.04	0.04	600
i	0.02	0.01	0.02	600
j	0.04	0.03	0.03	600
k	0.03	0.03	0.03	600
l	0.05	0.05	0.05	600
m	0.04	0.04	0.04	600
n	0.05	0.05	0.05	600
o	0.04	0.04	0.04	600
p	0.04	0.04	0.04	600
q	0.03	0.03	0.03	600
r	0.03	0.04	0.03	600
s	0.03	0.04	0.03	600
t	0.02	0.02	0.02	600
u	0.03	0.03	0.03	600
v	0.05	0.07	0.06	600
w	0.05	0.06	0.06	600
x	0.05	0.02	0.02	600
y	0.04	0.04	0.04	600
z	0.04	0.04	0.04	600
accuracy			0.04	15600
macro avg	0.04	0.04	0.04	15600
weighted avg	0.04	0.04	0.04	15600

Gambar 6. Hasil classification report tanpa horizontal flip

Dari hasil F1 Score diatas dapat dilihat rata - rata recall, precision, dan f1-score dari semua kelas adalah 0.04. Dari F1 score tersebut dapat dilihat bahwa model ini paling bagus memprediksi huruf v dan paling buruk saat memprediksi huruf i.

	precision	recall	f1-score	support
a	0.04	0.04	0.04	600
b	0.06	0.06	0.06	600
c	0.02	0.02	0.02	600
d	0.02	0.02	0.02	600
e	0.04	0.05	0.05	600
f	0.03	0.03	0.03	600
g	0.04	0.04	0.04	600
h	0.03	0.03	0.03	600
i	0.04	0.03	0.03	600
j	0.03	0.03	0.03	600
k	0.03	0.03	0.03	600
l	0.04	0.04	0.04	600
m	0.04	0.04	0.04	600
n	0.05	0.07	0.06	600
o	0.04	0.04	0.04	600
p	0.04	0.04	0.04	600
q	0.05	0.05	0.05	600
r	0.05	0.05	0.05	600
s	0.03	0.04	0.04	600
t	0.04	0.03	0.04	600
u	0.03	0.03	0.03	600
v	0.04	0.04	0.04	600
w	0.03	0.03	0.03	600
x	0.03	0.02	0.02	600
y	0.04	0.04	0.04	600
z	0.04	0.04	0.04	600
accuracy			0.04	15600
macro avg	0.04	0.04	0.04	15600
weighted avg	0.04	0.04	0.04	15600

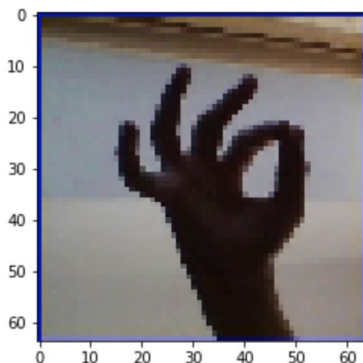
Gambar 7. Hasil classification report tanpa horizontal flip

Dari hasil F1 Score diatas dapat dilihat rata - rata recall, precision, dan f1-score dari semua kelas adalah 0.04. Dari F1 score tersebut dapat dilihat bahwa model ini paling bagus memprediksi huruf n dan paling buruk saat memprediksi huruf c, d, dan x.

F. Percobaan Model pada Gambar

file name: F_test.jpg

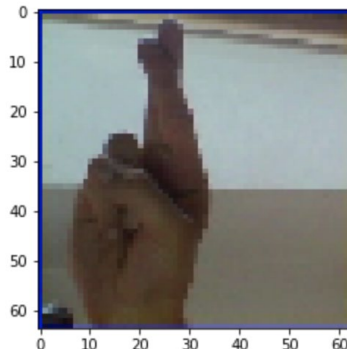
Hasil Prediksi: f



Gambar 8. Percobaan model pada gambar training 1

file name: R_test.jpg

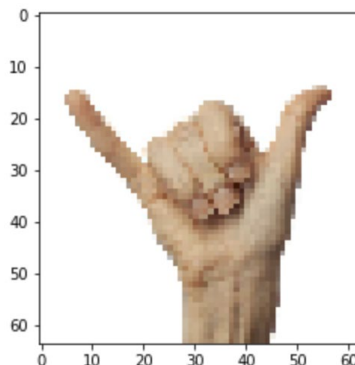
Hasil Prediksi: r



Gambar 9. Percobaan model pada gambar training 2

file name: y.jpeg

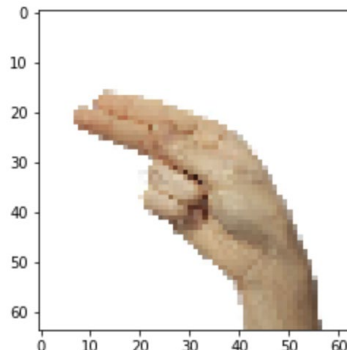
Hasil Prediksi: y



Gambar 10. Percobaan model pada gambar lain 1

file name: h.jpeg

Hasil Prediksi: p



Gambar 11. Percobaan model pada gambar lain 2

Dari percobaan diatas dapat dilihat test yang dilakukan pada dataset sendiri memiliki hasil prediksi yang benar sedangkan pada gambar lain tidak selalu benar. Seperti pada gambar 12 dapat dilihat bahwa gambar sebenarnya adalah h tetapi diprediksi sebagai p.

IV. KESIMPULAN

Pada penelitian ini dilakukan penelitian pada klasifikasi bahasa isyarat Amerika. Model CNN ini dibuat dengan menggunakan beberapa layer seperti *Conv2D*, *Dropout*, *Flatten*, dan *Dense*. Beberapa hal yang dilakukan untuk mengetahui perbedaan akurasi adalah pada *pre-processing data*, dan epoch. Dari penelitian ini hasil akurasi yang didapatkan adalah 82,1%. Dari penelitian ini ditemukan loss yang tinggi pada validation accuracy sedangkan loss pada training cukup stabil. Huruf yang diprediksi paling bagus adalah v dan n sedangkan paling buruk saat memprediksi huruf i, c, d, dan x. Untuk pengembangan kedepannya akan dilakukan beberapa hal seperti memperbaiki model dengan cara mengubah layernya dan juga melakukan preprocessing yang berbeda agar nilai loss dapat berkurang dan memberikan akurasi yang lebih tinggi.

UCAPAN TERIMA KASIH

(JIKA DIPERLUKAN)

Peneliti berterima kasih pada Bu Caecilia Citra Lestari S.Kom., M.Kom. dan Bapak Evan Tanuwijaya S.Kom, M.Kom. karena berkat bimbingan mereka peneliti dapat menyelesaikan jurnal. Peneliti juga berterima kasih pada orang tua dan keluarga penulis yang selalu mendukung dan menyemangati dari segala aspek mulai dari mental hingga keuangan. Tak lupa, peneliti berterima kasih pada teman - teman peneliti yang menyemangati dan selalu berjuang bersama dalam suka maupun duka

DAFTAR PUSTAKA

- [1] K. Kesehatan RI, "SITUASI PENYANDANG DISABILITAS", Pusdatin.kemkes.go.id, 2014. [Online]. Available: <https://pusdatin.kemkes.go.id/download.php?file=download/pusdatin/buletin/buletin-disabilitas.pdf>. [Accessed: 05- Jun- 2021]
- [2] "Sumbar Hanya Punya 8 Penerjemah Bahasa Isyarat |Republika Online", Republika Online, 2019. [Online]. Available: <https://www.republika.co.id/berita/pzjwqr335/sumbar-hanya-punya-8-penerjemah-bahasa-isyarat>. [Accessed: 05- Jun- 2021]
- [3] L. Zheng, B. Liang and A. Jiang, "Recent Advances of Deep Learning for Sign Language Recognition," 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2017, pp. 1-7, doi: 10.1109/DICTA.2017.8227483.
- [4] H. Kulhandjian, P. Sharma, M. Kulhandjian and C. D'Amours, "Sign Language Gesture Recognition Using Doppler Radar and Deep Learning," 2019 IEEE Globecom Workshops (GC Wkshps), 2019, pp. 1-6, doi: 10.1109/GCWkshps45667.2019.9024607.
- [5] M.A. Jalal, R. Chen, R.K. Moore and L. Mihaylova, "American sign language posture understanding with deep neural networks," in Proc. of IEEE International Conference on Information Fusion (FUSION), Cambridge, UK, Jul. 2018, pp. 573-579.
- [6] G. A. Rao, K. Syamala, P. V. V. Kishore and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), 2018, pp. 194-197, doi: 10.1109/SPACES.2018.8316344.
- [7] A. Das, S. Gawde, K. Suratwala and D. Kalbande, "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images," 2018 International Conference on Smart City and Emerging Technology (ICSCET), 2018, pp. 1-6, doi: 10.1109/ICSCET.2018.8537248.
- [8] "ASL Alphabet", Kaggle.com, 2018. [Online]. Available: <https://www.kaggle.com/grassknotted/asl-alphabet>. [Accessed: 10- Jun- 2021]
- [9] "tf.keras.preprocessing.image.ImageDataGenerator", TensorFlow, [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator. [Accessed: 05- Jun- 2021]
- [10] Suresh, S., T. P., M. H., & M.H, S. (2019). Sign Language Recognition System Using Deep Neural Network. 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). doi:10.1109/icaccs.2019.8728411
- [11] Bheda, Vivek & Radpour, Dianna. (2017). Using Deep Convolutional Networks for Gesture Recognition in American Sign Language.
- [12] K. Team, "Keras documentation: The Sequential model", Keras.io, [Online]. Available: https://keras.io/guides/sequential_model/. [Accessed: 05- Jun- 2021]
- [13] K. Team, "Keras documentation: Conv2D layer", Keras.io. [Online]. Available: https://keras.io/api/layers/convolution_layers/convolution2d/. [Accessed: 5- Jun- 2021]
- [14] K. Team, "Keras documentation: Dropout layer", Keras.io. [Online]. Available: https://keras.io/api/layers/regularization_layers/dropout/. [Accessed: 05- Jun- 2021]
- [15] K. Team, "Keras documentation: Flatten layer", Keras.io. [Online]. Available: https://keras.io/api/layers/reshaping_layers/flatten/. [Accessed: 10- Jun- 2021]