

# Deteksi Aktivitas Manusia Berdasarkan Data Skeleton dengan Menggunakan Modifikasi VGG16

Daniel Subroto, Liliana

Program Studi Informatika, Fakultas Teknologi Industri, Universitas Kristen Petra

Jln. Siwalankerto 121 – 131 Surabaya 60236

Telp. (031)-2983455, Fax. (031)-8417658

E-Mail: cxyren12@gmail.com, lilian@petra.ac.id

## ABSTRAK

Pada umumnya deteksi aktivitas manusia dilakukan untuk mendeteksi aktivitas pada kehidupan sehari-hari. Dengan pengembangan lebih lanjut, deteksi aktivitas manusia dimanfaatkan untuk mendeteksi aktivitas yang dicurigai (bukan rutinitas) sebagai aplikasi *early warning*. Deteksi aktivitas manusia dalam aplikasi *early warning* kemudian akan diimplementasikan pada sistem lainnya. Namun, terdapat beberapa permasalahan dalam mendeteksi aktivitas manusia antara lain, adanya variasi dalam melakukan suatu aktivitas, adanya gerakan transisi antar aktivitas, dan adanya kemiripan gerakan dalam aktivitas yang berbeda. Deteksi aktivitas manusia dengan variasi yang ada dapat dilakukan jika memanfaatkan pendekatan *deep learning* untuk melakukan proses training. Metode *deep learning* yang digunakan adalah VGG16.

VGG16 akan menerima *input* yang berupa gambar data skeleton. Data skeleton yang digunakan diperoleh dari dataset NTU RGB+D. Data skeleton akan direpresentasikan sebagai gambar 2D dengan melalui proses meliputi *crop*, diubah menjadi *grayscale*, *resize*, dan mengabungkan 10 gambar menjadi 10 *channels* untuk setiap rangkaian urutan aktivitas. Untuk mendeteksi aktivitas manusia diterapkan *transfer learning* pada VGG16 yaitu mengubah *fully connected layer*.

Hasil pengujian modifikasi VGG16 dengan representasi data skeleton menghasilkan tingkat akurasi tertinggi sebesar 54.59%. Tingkat akurasi ini didapatkan dari pengujian model menggunakan dataset yang sama dengan dataset *training*. Modifikasi VGG16 masih menjadi model yang terbaik berdasarkan pengujian dengan model *Convolutional Neural Network* lainnya. Modifikasi VGG16 dapat mengklasifikasi aktivitas dalam ruangan.

**Kata Kunci:** Deteksi Aktivitas Manusia, NTU RGB+D, *Convolutional Neural Network*, VGG16, Data Skeleton

## ABSTRACT

*In general, detection of human activity is carried out to detect activity in daily life. With further development, the detection of human activity is utilized to detect suspected activity (not routine) as an early warning application. Detection of human activity in early warning applications will then be implemented in other systems. However, there are several problems in detecting human activity, among others, the presence of variations in performing an activity, the movement of transitions between activities, and the similarity of movements in different activities. Detection of human activities with existing variations can be done if utilizing a deep learning approach to conduct the training process. The deep learning method used is VGG16.*

*VGG16 will receive input in the form of skeleton data images. The skeleton data used is obtained from the NTU RGB+D dataset. Skeleton data will be represented as 2D images by going through a process of covering the crop, converted into grayscale, resizing, and connecting 10 images into 10 channels for each sequence of activity sequences. To detect human activity is applied transfer learning on VGG16 that is changing the fully connected layer.*

*VGG16 modification test results with skeleton data representation resulting in the highest accuracy rate of 54.59%. This level of accuracy is obtained from model testing using the same dataset as the training dataset. The VGG16 modification is still the best model based on testing with other Convolutional Neural Network models. Modification of VGG16 can classify indoor activities.*

**Keywords:** Human Activity Recognition, NTU RGB+D, Convolutional Neural Network, VGG16, Skeleton data.

## 1. PENDAHULUAN

Deteksi aktivitas manusia merupakan salah satu topik penelitian yang mutakhir karena ketersediaan sensor yang memadai, konsumsi daya yang rendah dan murah; kemajuan *computer vision*; *machine learning*; kecerdasan buatan; dan IoT [2, 3, 12, 18, 20, 21]. Deteksi aktivitas manusia dapat dilakukan menggunakan data yang didapat dari sensor *wearable* atau kamera. Data dari sensor *wearable* berupa data *accelerometer*, data *gyroscope*, data *acceleration signals*, dan data *whirling signals*. Sedangkan data dari kamera berupa video atau gambar. Data dari video dapat diekstrak menjadi data *frame* gambar atau data skeleton. Dengan pengembangan lebih lanjut deteksi aktivitas manusia dapat digunakan untuk aplikasi *early warning*. Deteksi aktivitas manusia dalam aplikasi *early warning* kemudian akan diimplementasikan pada sistem lainnya [3, 5, 8–11, 16, 17, 19, 23, 25–27]. Deteksi aktivitas manusia dapat digunakan untuk pemantauan via CCTV, sistem *smart home*, dan deteksi aktivitas manusia didalam ruangan. Deteksi aktivitas manusia dapat dilakukan secara *offline* dan *online* [14]. *Offline* adalah mendeteksi aktivitas manusia dengan data yang sudah ada lalu dideteksi aktivitas tersebut, sedangkan *online* adalah berarti mendeteksi aktivitas manusia secara *real time*.

Terdapat beberapa kendala dalam mendeteksi aktivitas manusia antara lain, adanya variasi dalam melakukan suatu aktivitas, adanya gerakan transisi antar aktivitas, dan adanya kemiripan gerakan dalam aktivitas yang berbeda. Kendala variasi terjadi karena dalam deteksi aktivitas manusia akan dipengaruhi oleh kepribadiannya dan keadaan psikologisnya sehingga informasi tersebut sulit untuk diekstrak [1, 11, 24]. Kendala berikutnya adalah aktivitas terdiri dari untaian gerakan yang lebih sederhana yang membentuk rangkaian urutan gerakan. Selain itu, kesamaan

*intra*class dan *inter*class membuat masalahnya sangat menantang [13, 24]. Artinya, aktivitas dalam kelas yang sama dapat diungkapkan oleh orang yang berbeda dengan gerakan tubuh yang berbeda dan aktivitas antar kelas yang berbeda mungkin akan sulit dibedakan serta sebaliknya karena terdapat informasi yang mirip. Cara seseorang melakukan aktivitas tergantung pada kebiasaannya dan hal ini yang membuat masalah dalam mengidentifikasi aktivitas yang mendasarinya akan sulit untuk ditentukan. Kendala-kendala yang muncul dalam mendeteksi aktivitas manusia dapat diatasi jika proses *training* dilakukan dengan menggunakan pendekatan *machine learning*. Namun demikian, penggunaan metode *machine learning* sederhana kurang bisa menghasilkan hasil yang memuaskan [7]. Oleh karena itu, perlu digunakannya *deep learning* dalam mendeteksi aktivitas manusia. Kemampuan *deep learning* dalam mengekstrak fitur tingkat tinggi membuat *deep learning* lebih unggul dibandingkan dengan menggunakan *machine learning* konvensional.

Terdapat berbagai penelitian mengenai deteksi aktivitas manusia yang pernah dilakukan sebelumnya. Dari berbagai penelitian terdapat penggunaan dataset yang berbeda-beda dalam mendeteksi aktivitas manusia. Contohnya adalah menggunakan data sensor dari *smartphone* [23], menggunakan data skeleton [6], dan lain-lainnya. Berdasarkan dari penelitian tersebut, data skeletonlah yang bisa mendekati hasil mendeteksi aktivitas manusia yang diinginkan. Selain itu, terdapat berbagai metode *deep learning* yang dapat digunakan untuk mendeteksi aktivitas manusia. Contohnya adalah menggunakan *Convolutional Neural Network* [6], *Support Vector Machine* [4], dan lain-lainnya.

Oleh karena itu digunakannya data skeleton dan metode *Convolutional Neural Network*. Representasi data skeleton akan menggunakan data gambar sehingga dapat diolah oleh CNN dan dapat menghasilkan tingkat akurasi cukup tinggi. Pada penelitian ini proses *training* untuk deteksi aktivitas manusia dilakukan dengan menggunakan modifikasi VGG16 berdasarkan data skeleton. Aktivitas yang akan dideteksi adalah aktivitas yang bisa dilakukan dalam keseharian. Kemudian aktivitas-aktivitas tersebut akan diproses dan dilakukan klasifikasi dengan menggunakan modifikasi VGG16. Proses klasifikasi dilakukan berdasarkan data skeleton yang berbentuk kumpulan data sendi. Data skeleton didapatkan dari perangkat Microsoft Kinect versi 2. Model yang dibuat akan dapat melakukan klasifikasi berdasarkan data skeleton yang akan diubah menjadi bentuk gambar. Kemudian gambar data skeleton akan dilakukan klasifikasi, dengan menggunakan data skeleton proses klasifikasi aktivitas akan lebih maksimal dan lebih optimal dibandingkan dengan menggunakan data video RGB biasa. Model yang dibuat tidak hanya mendeteksi pose gerakan saja, tetapi akan mendeteksi dari kumpulan pose gerakan. Dengan demikian, model akan mendapatkan informasi yang sesuai dengan gerakan aktivitas.

## 2. TINJAUAN STUDI

Subasi et al. [23], melakukan penelitian untuk mendeteksi aktivitas manusia dengan menggunakan data sensor *wearable* pada perangkat *smartphone*. Pada penelitian ini menggunakan berbagai metode yaitu, *Artificial Neural Network*, *k-Nearest Neighbor*, *Support Vector Machine*, *Naïve Bayes*, *Classification and Regression Tree*, *C4.5 Decision Tree*, *REPTree*, *LADTree Algorithm*, dan *Random Forest*. Hasil yang didapatkan cukup memuaskan dengan hasil terbaik oleh metode *Support Vector Machine* dengan tingkat akurasi sebesar 98.91%. Percobaan dilakukan dengan menggunakan 5 jenis aktivitas. Kelemahan dari penggunaan data sensor *wearable* adalah data sensor hanya

terbatas, apabila *smartphone* diletakkan maka tidak dapat mendeteksi aktivitas manusia.

Du et al. [6], mendeteksi aktivitas manusia berdasarkan data skeleton dengan menggunakan metode *Convolutional Neural Netork*. Data skeleton dalam penelitian ini akan digunakan dengan cara mengubah representasi data skeleton menjadi rangkaian urutan gambar. Metode *Convolutional Neural Network* yang diusulkan dapat menghasilkan tingkat akurasi sebesar 100%. Namun dengan menggunakan metode *Convolutional Neural Network* sulit memungkinkan untuk dapat menghasilkan tingkat akurasi sebesar 100%.

Cippitelli et al. [4], mendeteksi aktivitas manusia dengan menggunakan metode *Support Vector Machine* dan data skeleton. Representasi data skeleton yang digunakan untuk penelitian ini adalah dengan menggunakan data histogram dari data skeleton. Metode *Support Vector Machine* dapat menghasilkan tingkat akurasi sebesar 48.9% pada pengujian *cross-subject* dan 57.7% pada pengujian *cross-view*. Pada penelitian ini tingkat akurasi yang bisa didapatkan tidak setinggi penelitian-penelitian sebelumnya.

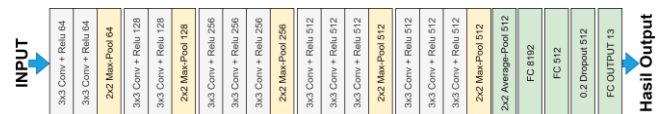
## 3. METODOLOGI

### 3.1 VGG16

VGG16 adalah kepanjangan dari *Visual Geometric Group* yang dikembangkan oleh Oxford University dan VGG16 memiliki arti yaitu terdapat 16layer yang terdapat dalam *Visual Geometric Group* [22]. Secara singkat VGG16 adalah salah satu konfigurasi arsitektur dalam *Convolutional Neural Network*.

VGG16 memiliki *input* layer yang berukuran 224x224. VGG16 terdapat 16layer *convolution* dengan *kernel* sebesar 3x3, *stride* sebesar 1, dan ReLu untuk setiap *hidden layer*. *Convolution layer* akan diikuti dengan 5 *Max Pooling layer*. *Max Pooling layer* yang digunakan memiliki *kernel* sebesar 2x2 dan *stride* sebesar 2. Kemudian akan diikuti dengan 3 *Fully Connected layer* yang 2 diantaranya memiliki 4096 *channels* dan *layer* terakhir menggunakan *softmax activation* untuk proses klasifikasi dengan jumlah *channels* yang sesuai dengan jumlah kelas.

Modifikasi VGG16 menerapkan konsep *transfer learning*. Model modifikasi VGG16 mengambil *pre-trained weights* dari model yang telah di-*train* sebelumnya. Kemudian model dari VGG16 ini akan dimodifikasi pada bagian *fully connected layer* seperti pada Gambar 1. Modifikasi ini meliputi menambahkan *Average Pooling Layer* dan *dropout*.



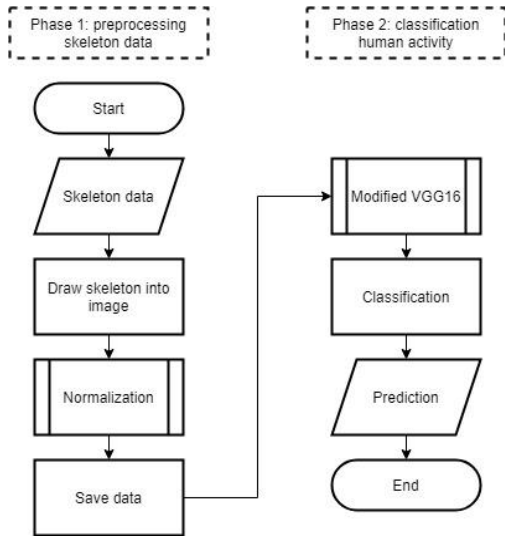
Gambar 1. Arsitektur modifikasi VGG16

Pada model modifikasi VGG16 harus melalui tahap *training* sebelum digunakan untuk melakukan klasifikasi aktivitas dengan data *testing*. Setelah dilakukan proses *training*, model akan melakukan klasifikasi pada data yang sudah disiapkan untuk *testing*.

### 3.2 Desain Sistem

Pada bagian ini membahas permasalahan bagaimana data yang digunakan sebagai *input* akan diproses oleh program sehingga dapat menghasilkan *output* yang sesuai. Penggambaran sistem seperti pada Gambar 2, sistem mencakup *preprocessing* data

skeleton, model *neural network* untuk *training*, dan sistem untuk *testing*. Sistem akan menerima *input* data dari user berupa data skeleton dalam format *.csv* atau *.skeleton*. Karena *input* dari modifikasi VGG16 membutuhkan gambar, maka data skeleton akan digambar ulang. Data gambar kemudian akan dilakukan proses normalisasi, kemudian gambar skeleton akan disimpan. Data tersebut akan di-olah dengan menggunakan modifikasi VGG16 yang telah di-*train* sebelumnya untuk dilakukannya klasifikasi aktivitas/gerakan manusia yang dilakukan oleh subjek dalam data skeleton tersebut. Hasil klasifikasi akan dicetak dalam bentuk tipe data *string*.



Gambar 2. Blok diagram sistem.

### 3.2.1 Dataset

Dataset NTU RGB+D yang digunakan memiliki resolusi sebesar 1920x1080 dan memiliki frame rate sebesar 30frame per detiknya. Dataset ini memiliki kelas aktivitas berjumlah 60 buah, dengan 40 subjek, perekaman dengan 3 kamera, dan 2 kali repetisi. Total datanya sejumlah 56880 data tetapi terdapat 302 data skeleton yang hilang. Apabila semua data diubah menjadi gambar dengan *mean* 90 frame maka data yang terkumpul adalah 5092020 gambar, hal ini tidak memastikan model dapat mendeteksi seluruh aktivitas manusia, karena ciri khas dari dataset ini adalah awalan posisi yang sama. Serta, beberapa kelas aktivitas tidak relevan. Oleh karena itu adanya pengurangan jumlah frame, pemilihan kelas, dan penggunaan konfigurasi kamera yang mirip dengan yang digunakan pada perekaman data untuk pengujian. Pengurangan jumlah frame dilakukan dengan mengambil 10 gambar dari video sebagai sample, pengambilan gambar dengan proporsi yang sama berdasarkan jumlah frame dalam data skeleton.

### 3.2.2 Preprocessing

Pada *preprocessing*, data skeleton akan diubah menjadi gambar dan kemudian akan dinormalisasi sehingga gambar siap untuk menjadi *input*. Untuk mencapai hasil yang terbaik penggambaran data skeleton akan menggunakan rumus garis *brensenham* sehingga garis akan sesuai dengan pixel yang ada. Kemudian gambar akan dinormalisasi. Proses normalisasi mengalami beberapa tahap untuk mencapai hasil maksimal.

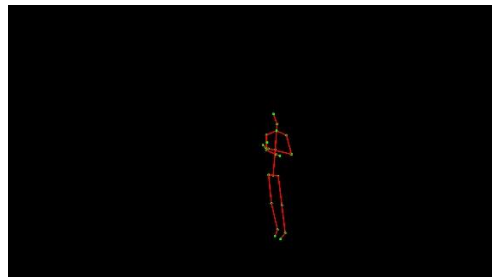
Pada awalnya data skeleton digambar menggunakan *background* video RGB, gambar dapat dilihat pada Gambar 3. Tetapi konfigurasi *background* pada dataset ini berada pada kantor,

sehingga *background* tidak sesuai dengan *background* pada situasi di rumah. Pengaruh *background* bisa mengurangi kemampuan modifikasi VGG16 dalam mengambil *features* yang penting.



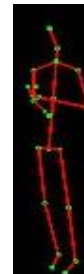
Gambar 3. Gambar data skeleton dengan *background* video RGB

Solusi alternatif untuk meningkatkan kemampuan modifikasi VGG16 adalah menghitamkan *background*, gambar dapat dilihat pada Gambar 4. Tetapi *background* hitam yang menyeluruh dan mendominasi akan mengakibatkan model mengalami kesusahan dalam membedakan aktivitas satu dengan lainnya.



Gambar 4. Gambar data skeleton dengan *background* hitam

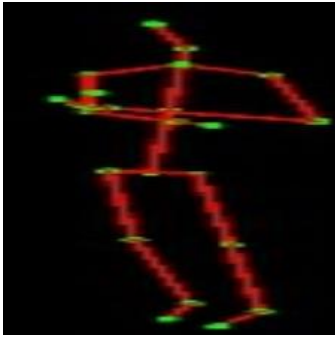
Kekurangan ini bisa diatasi dengan memotong pada bagian skeleton sehingga proporsi *background* menjadi lebih kecil seperti pada Gambar 5.



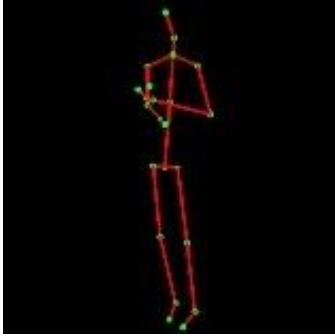
Gambar 5. Gambar data skeleton yang sudah dipotong

Jumlah dan ukuran data yang besar akan menghasilkan proses *training* menjadi lebih lama dan adanya ukuran yang tidak seragam dapat mengakibatkan model mengalami kesulitan dalam menyesuaikan *input*. Melakukan *resize* gambar akan memudahkan model dalam menyesuaikan *input*. Dengan demikian semua ukuran akan seragam, namun apabila gambar dilakukan *resize* secara langsung gambar akan mengalami *stretch*, gambar dapat dilihat pada Gambar 6.

Solusi agar gambar tidak mengalami *stretch* adalah melakukan *mapping* ulang gambar seperti pada Gambar 7. Selain itu terdapat perbedaan orientasi gambar, aktivitas yang berbentuk horizontal dan vertikal sehingga *mapping* perlu dilakukan. *Mapping* dilakukan dengan cara meletakkan gambar di tengah gambar kosong yang ukurannya sudah di-*resize*.

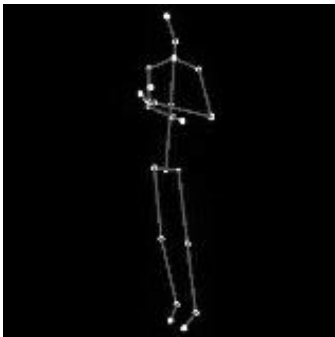


Gambar 6. Gambar data skeleton yang mengalami *stretch*



Gambar 7. Gambar data skeleton yang sudah dilakukan *mapping* dan *resize*.

Untuk mendeteksi aktivitas manusia, gambar skeleton yang sudah dilakukan *resize* akan digunakan sebagai data *input*. Gambar skeleton berbentuk RGB memiliki 3 *channels* tetapi data skelon tidak berpengaruh dengan warna sehingga data skeleton dalam bentuk *grayscale* akan mempermudah belajar model dan melakukan klasifikasi. Gambar *grayscale* dijadikan biner menjadi 0 dan 255 sehingga modifikasi VGG16 akan mendapatkan *feature* yang lebih optimal. Gambar *grayscale* dapat dilihat pada Gambar 8.

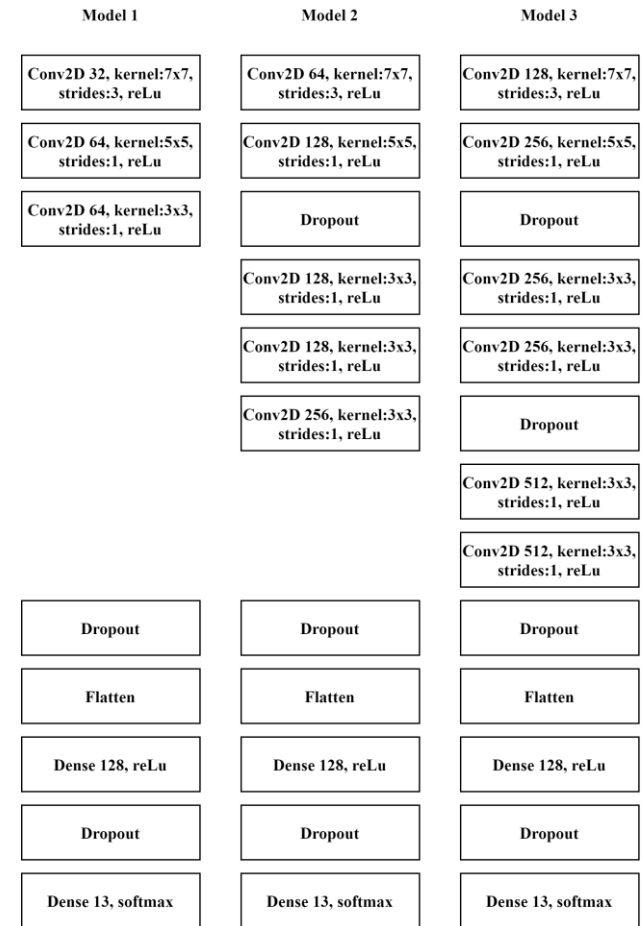


Gambar 8. Gambar *grayscale* yang sudah dinormalisasi.

Gambar yang sudah menjadi *grayscale* akan dilakukan *training*, namun pada proses *training* gambar yang digunakan adalah satu persatu sehingga model akan melakukan klasifikasi secara satu persatu. Sehingga proses klasifikasi tidak akan bisa mendapatkan hasil yang pasti karena hasil prediksi perframe akan berbeda dan juga gerakan perlu dilakukan secara berurutan sedangkan dalam modifikasi VGG16 tidak bisa mendeteksi secara berurutan. Solusi untuk mengatasi hal ini adalah dengan melakukan *channel* 10 gambar secara bersamaan. Sehingga *input* yang didapatkan modifikasi VGG16 adalah 10 gambar yang bergerak sesuai dengan gerakan yang digunakan.

### 3.2.3 Model Lain

Terdapat beberapa arsitektur model CNN yang digunakan, salah satunya adalah model VGG16 yang telah dimodifikasi, model usulan dari penelitian [6], dan 3 buah model yang bereferensi dari [15]. Model [6] dan 3 model lainnya akan digunakan untuk menguji model modifikasi VGG16. Arsitektur 3 model lainnya dapat dilihat pada Gambar 9.



Gambar 9. Arsitektur 3 model lainnya, 3 model ini bereferensi dari [15]

## 4. PENGUJIAN

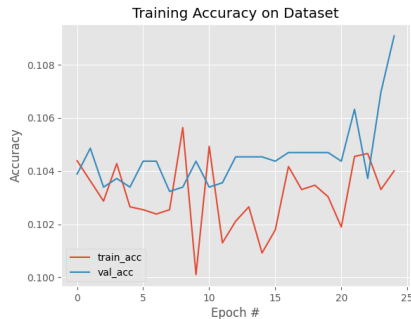
Pada bagian ini membahas tentang pengujian tingkat akurasi terhadap model yang dibuat. Pengujian dilakukan dengan menggunakan 13 kelas aktivitas yang sudah ditentukan sebelumnya. Pengujian model dilakukan dengan mencari konfigurasi *preprocessing* yang tepat yaitu pengujian data skeleton dengan cara *image preprocessing*. Kemudian akan menguji model modifikasi VGG16 dengan model CNN lainnya.

### 4.1 Pengujian Data Skeleton

Pengujian data skeleton bertujuan untuk mencari pendekatan yang paling tepat dalam pengolahan data skeleton sehingga model *Convolutional Neural Network* dapat menghasilkan akurasi yang tinggi. Pengujian ini meliputi 5 representasi dataset yang berbeda. 5 representasi tersebut adalah gambar skeleton dengan *background* video RGB, gambar skeleton dengan *background* hitam, gambar skeleton yang telah dipotong, gambar skeleton dengan 30 *channels*, dan gambar skeleton dengan 10 *channels*.

#### 4.1.1 Gambar Skeleton dengan background video RGB

Pengujian ini dilakukan dengan menggunakan gambar skeleton dengan *background* video RGB. Adapun pengujian dilakukan dengan *transfer learning* model modifikasi VGG16 dengan konfigurasi *batchsize* 10, *learning rate* 0.001, *epoch* 25, dan *dropout* 0.5.

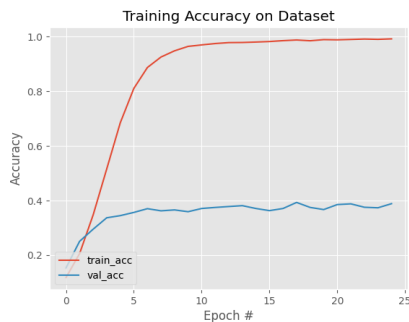


**Gambar 10** Hasil tingkat akurasi *training* dan *validasi* per *epoch* dengan menggunakan gambar skeleton dengan *background* video RGB

Gambar 10. menunjukkan bahwa penggunaan gambar data skeleton dengan *background* video RGB tidak dapat menghasilkan akurasi yang tinggi. Penyebab akurasi yang rendah adalah karena terlalu banyak *background* yang tidak diperlukan sehingga kemampuan model untuk mendapatkan *feature* skeletonnya berkurang.

#### 4.1.2 Gambar Skeleton dengan background hitam

Pengujian ini dilakukan dengan menggunakan data skeleton dengan *background* hitam. Pengujian ini digunakan sebagai dataset untuk melakukan konfigurasi model modifikasi VGG16.

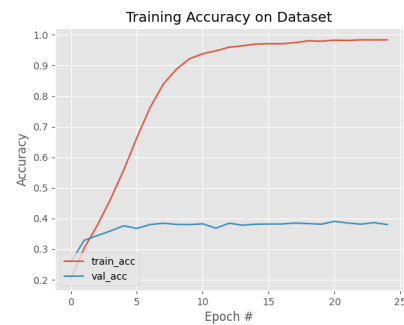


**Gambar 11.** Hasil tingkat akurasi *training* dan *validasi* per *epoch* dengan menggunakan gambar skeleton dengan *background* Hitam

Gambar 11. menunjukkan bahwa penggunaan gambar data skeleton dengan *background* hitam dapat menghasilkan akurasi yang lebih tinggi dibandingkan dengan representasi data sebelumnya. Adanya peningkatan tingkat akurasi dikarenakan oleh *background* yang telah berubah, sebelumnya dengan *background* video RGB model sulit untuk mendapatkan data skeletonnya sedangkan dengan *background* hitam model akan lebih mudah mendapatkan data skeletonnya. Namun pengambilan data skeleton masih belum maksimal dikarenakan *background* masih mendominasi dibandingkan dengan data skeleton yang ada.

#### 4.1.3 Gambar Skeleton yang telah dipotong

Setelah melihat hasil pengujian sebelumnya, penggunaan data skeleton dengan *background* hitam tidak bisa menghasilkan tingkat akurasi yang maksimal. Oleh karena itu data gambar skeleton akan dipotong sehingga diharapkan dapat menaikkan tingkat akurasi pada model. Penggunaan gambar data skeleton yang telah dipotong dilakukan beberapa kali percobaan. Namun gambar data skeleton tidak melakukan *mapping* ulang sehingga gambar mengalami *stretch*. Percobaan-percobaan tersebut tidak menghasilkan akurasi yang tinggi walaupun sudah menggunakan konfigurasi yang terbaik dari model modifikasi VGG16.

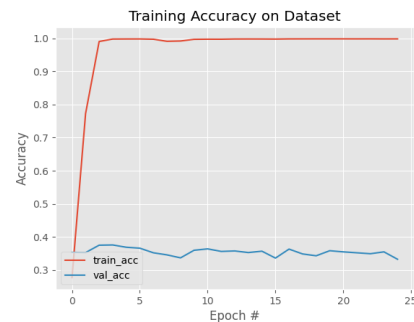


**Gambar 12.** Hasil tingkat akurasi *training* dan *validasi* per *epoch* dengan menggunakan gambar skeleton yang telah dipotong

Gambar 12. menunjukkan bahwa penggunaan data skeleton yang telah dipotong tidak mendapatkan peningkatan tingkat akurasi yang signifikan. Peningkatan yang tidak signifikan ini terjadi karena gambar skeleton yang digunakan mengalami *stretch* sehingga *feature* mengalami perubahan dan model akan lebih sulit untuk melakukan klasifikasi antara kelas yang digunakan. Apabila hanya digunakan satu persatu gambar akan sulit dalam membedakan antar aktivitas karena aktivitas dilakukan dengan berbagai gerakan dalam suatu urutan sehingga perlu dilakukan proses transisi yang urut antar gerakan dalam suatu aktivitas.

#### 4.1.4 Gambar Skeleton dengan 30 Channels

Pengujian ini dilakukan dengan menggunakan gambar yang sudah dipotong kemudian dilakukan *mapping* ulang, bentuk RGB tetap dipertahankan, dan menggabungkan 10 gambar menjadi 1 kesatuan. Karena bentuk RGB tetap dipertahankan maka apabila 10 gambar digabung menjadi 1, maka yang terbentuk adalah data dengan 30 *channels*.



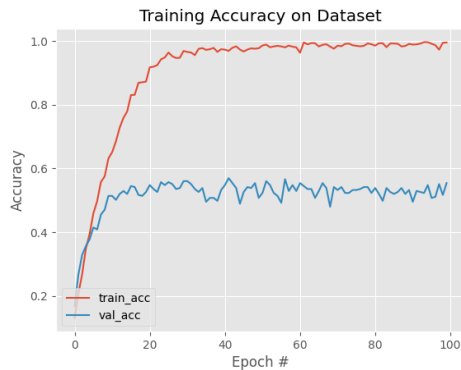
**Gambar 13.** Hasil tingkat akurasi *training* dan *validasi* per *epoch* dengan menggunakan gambar skeleton dengan 30 *channels*



Gambar 13. menunjukkan bahwa penggunaan gambar data skeleton dengan 30 *channels* tidak menghasilkan hasil yang baik. Dengan tetap mempertahankan gambar RGB maka terdapat beberapa *channel* gambar yang kosong dan *channel* gambar yang hanya terdapat garis atau titik sendi saja. Sedangkan data yang dibutuhkan adalah data secara menyeluruh yaitu seluruh skeleton dalam suatu *channel* gambar.

#### 4.1.5 Gambar Skeleton dengan 10 Channels

Pengujian ini dilakukan dengan menggunakan gambar yang sudah dipotong kemudian dilakukan *mapping* ulang, mengubahnya menjadi *grayscale* dan menjadikannya biner, dan menjadikan 10 *channels* gambar.



**Gambar 14. Hasil tingkat akurasi *training* dan *validasi* per *epoch* dengan menggunakan gambar skeleton dengan 10 *channels***

Gambar 14. menunjukkan bahwa penggunaan gambar data skeleton dengan 10 *channel* merupakan hasil terbaik dibandingkan dengan representasi data skeleton lainnya. Hasil ini didapatkan bahwa tingkat akurasi *training* bisa mencapai sebesar 99.72%, sedangkan untuk tingkat akurasi *validasi* bisa mencapai sebesar 56.96%.

## 4.2 Pengujian Model

Pengujian tingkat akurasi berdasarkan hasil uji tiap model. Pengujian ini dilakukan dengan menggunakan representasi dataset yang terakhir, yaitu gambar skeleton dengan 10 *channels*. Pengujian ini dapat mengetahui pengaruh model dengan dataset dengan cara melakukan pengujian dengan model yang berbeda.

**Table 1. Hasil tingkat akurasi tiap model**

Model	Akurasi <i>testing</i>	Mean akurasi <i>training</i>	Mean akurasi <i>validasi</i>	Waktu <i>training</i> (detik)
Model 1	32.26%	96.35%	27.77%	2769
Model 2	36.35%	64.46%	31.58%	7207
Model 3	35.48%	30.77%	25.18%	7307
[6]	37.34%	94.34%	31.87%	1865
Modifikasi VGG16	54.59%	90.22%	51.57%	12071

Table 1. menunjukkan bahwa modifikasi VGG16 memiliki hasil lebih baik dibanding dengan model lainnya. Modifikasi VGG16 mendapatkan hasil yang terbaik walaupun waktu untuk *training* cukup lama. Pengaruh jumlah kedalaman, jumlah filter, adanya *dropout* saling mempengaruhi tingkat akurasi model. Jika jumlah layer terlalu sedikit maka data yang dilatih tidak akan bisa maksimal dalam mendapatkan *features*.

## 5. KESIMPULAN

Dari hasil perancangan dan pembuatan sistem dan aplikasi, dapat disimpulkan bahwa:

- Penggunaan CNN dengan representasi data skeleton dalam bentuk gambar dirasa tidak tepat, karena representasi data skeleton harus saling berhubungan menjadi suatu urutan transisi pose gerakan namun CNN kurang bisa memaksimalkan hal tersebut. Pendekatan yang terbaik adalah dengan menggunakan gambar data skeleton dengan 10 *channels*.
- Dari hasil uji coba yang telah dilakukan diketahui bahwa pengolahan data skeleton yang baik dapat menghasilkan hasil yang maksimal, data skeleton yang baik adalah yang tidak terdistorsi informasi *background* yang lebih dominan dibandingkan dengan informasi skeletonnya. Pengolahan data skeleton yang baik akan sangat berpengaruh dengan model yang akan dibuat. Apabila data yang digunakan terlalu banyak *background* atau *noise* akan mengakibatkan metode CNN tidak bisa bekerja secara maksimal.
- Model yang terbaik adalah model yang diadaptasi dari model modifikasi VGG16 apabila dibandingkan dengan model lain dengan dataset yang sama. Model modifikasi VGG16 dalam training dapat menghasilkan tingkat akurasi rata-rata sebesar 90.22%. Sedangkan tingkat akurasi selama testing bisa menghasilkan tingkat akurasi sebesar 54.59%.

## 6. SARAN

Saran yang diberikan untuk pengembangan kedepannya:

- Mengkombinasi modifikasi VGG16 dengan metode lain atau mengganti modifikasi VGG16 dengan metode lain.
- Menggunakan data skeleton dengan representasi 3D data, tidak menggunakan format gambar.
- Melakukan data *augmentation* sehingga mendapatkan data yang lebih bervariasi dan mendapatkan akurasi yang lebih tinggi.

## 7. REFERENCES

- [1] Almaadeed, N., Elharrouss, O., Al-Maadeed, S., Bouridane, A. and Beghdadi, A. 2019. A Novel Approach for Robust Multi Human Action Detection and Recognition based on 3-Dimensional Convolutional Neural Networks. (2019), 1–7.
- [2] Bevilacqua, A., MacDonald, K., Rangarej, A., Widjaya, V., Caulfield, B. and Kechadi, T. 2019. *Human Activity Recognition with Convolutional Neural Networks*. Springer International Publishing.
- [3] Browne, D., Michael, G. and Steven, P. 2019. Deep learning human activity recognition. *CEUR Workshop Proceedings*. 2563, (2019), 76–87. DOI: 10.1016/j.neucom.2020.11.020.
- [4] Cippitelli, E., Gambi, E., Spinsante, S. and Flórez-Revuelta, F. 2016. Evaluation of a skeleton-based method for human activity recognition on a large-scale RGB-D dataset. *IET Conference Publications*. 2016, 4 (2016). DOI: 10.1049/ic.2016.0063.
- [5] Dhiman, C., Saxena, M. and Vishwakarma, D.K. 2019.

- Skeleton-based view invariant deep features for human activity recognition. *Proceedings - 2019 IEEE 5th International Conference on Multimedia Big Data, BigMM 2019*. (2019), 225–230. DOI:10.1109/BigMM.2019.00-21.
- [6] Du, Y., Fu, Y. and Wang, L. 2016. Skeleton based action recognition with convolutional neural network. *Proceedings - 3rd IAPR Asian Conference on Pattern Recognition, ACPR 2015*. (2016), 579–583. DOI: 10.1109/ACPR.2015.7486569.
- [7] Feng, S. and Duarte, M.F. 2019. Few-shot learning-based human activity recognition. *Expert Systems with Applications*. 138, (2019), 112782. DOI: 10.1016/j.eswa.2019.06.070.
- [8] Franco, A., Magnani, A. and Maio, D. 2020. A multimodal approach for human activity recognition based on skeleton and RGB data. *Pattern Recognition Letters*. 131, (2020), 293–299. DOI: 10.1016/j.patrec.2020.01.010.
- [9] Hussain, Z., Sheng, Q.Z. and Zhang, W.E. 2020. A review and categorization of techniques on device-free human activity recognition. *Journal of Network and Computer Applications*. 167, June (2020), 102738. DOI: 10.1016/j.jnca.2020.102738.
- [10] Ignatov, A. 2018. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing Journal*. 62, (2018), 915–922. DOI:10.1016/j.asoc.2017.09.027.
- [11] Janani, M., Nataraj, M. and Ganesh, C.R.S. 2020. *Mining and monitoring human activity patterns in smart environment-based healthcare systems*. Elsevier Inc.
- [12] Jobanputra, C., Bavishi, J. and Doshi, N. 2019. Human activity recognition: A survey. *Procedia Computer Science*. 155, 2018 (2019), 698–703. DOI: 10.1016/j.procs.2019.08.100.
- [13] Lee, I., Kim, D., Kang, S. and Lee, S. 2017. Ensemble Deep Learning for Skeleton-Based Action Recognition Using Temporal Sliding LSTM Networks. *Proceedings of the IEEE International Conference on Computer Vision*. 2017-Octob, (2017), 1012–1020. DOI: 10.1109/ICCV.2017.115.
- [14] Li, Y., Lan, C., Xing, J., Zeng, W., Yuan, C. and Liu, J. 2016. Online human action detection using joint classification-regression recurrent neural networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 9911 LNCS, (2016), 203–220. DOI: 10.1007/978-3-319-46478-7\_13.
- [15] Liliana, Chae, J.H., Lee, J.J. and Lee, B.G. 2020. A robust method for VR-based hand gesture recognition using density-based CNN. *Telkommnika (Telecommunication Computing Electronics and Control)*. (2020). DOI:10.12928/TELKOMNIKA.v18i2.14747.
- [16] Liu, J., Shahroudy, A., Wang, G., Duan, L.Y. and Kot, A.C. 2020. Skeleton-Based Online Action Prediction Using Scale Selection Network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 42, 6 (2020), 1453–1467. DOI: 10.1109/TPAMI.2019.2898954.
- [17] Núñez, J.C., Cabido, R., Pantrigo, J.J., Montemayor, A.S. and Vélez, J.F. 2018. Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition. *Pattern Recognition*. 76, (2018), 80–94. DOI: 10.1016/j.patcog.2017.10.033.
- [18] Seemanthini, K. and Manjunath, S.S. 2018. Human Detection and Tracking using HOG for Action Recognition. *Procedia Computer Science*. 132, Iccids (2018), 1317–1326. DOI: 10.1016/j.procs.2018.05.048.
- [19] Si, C., Chen, W., Wang, W., Wang, L. and Tan, T. 2019. An attention enhanced graph convolutional lstm network for skeleton-based action recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2019-June, (2019), 1227–1236. DOI: 10.1109/CVPR.2019.00132.
- [20] Si, C., Jing, Y., Wang, W., Wang, L. and Tan, T. 2020. Skeleton-based action recognition with hierarchical spatial reasoning and temporal stack learning network. *Pattern Recognition*. 107, xxxx (2020). DOI: 10.1016/j.patcog.2020.107511.
- [21] Si, C., Jing, Y., Wang, W., Wang, L. and Tan, T. 2018. Skeleton-Based Action Recognition with Spatial Reasoning and Temporal Stack Learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 11205 LNCS, (2018), 106–121. DOI: 10.1007/978-3-030-01246-5\_7.
- [22] Simonyan, K. and Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. (2015), 1–14.
- [23] Subasi, A., Khateeb, K., Brahimi, T. and Sarirete, A. 2020. *Human activity recognition using machine learning methods in a smart healthcare environment*. Elsevier Inc.
- [24] Vrigkas, M., Nikou, C. and Kakadiaris, I.A. 2015. A review of human activity recognition methods. *Frontiers Robotics AI*. 2, NOV (2015), 1–28. DOI: 10.3389/frobt.2015.00028.
- [25] Yan, S., Li, Z., Xiong, Y., Yan, H. and Lin, D. 2019. Convolutional sequence generation for skeleton-based action synthesis. *Proceedings of the IEEE International Conference on Computer Vision*. 2019-Octob, (2019), 4393–4401. DOI: 10.1109/ICCV.2019.00449.
- [26] Zhou, K., Wu, T., Wang, C., Wang, J. and Li, C. 2020. Skeleton Based Abnormal Behavior Recognition Using Spatio-Temporal Convolution and Attention-Based LSTM. *Procedia Computer Science*. 174, 2019 (2020), 424–432. DOI: 10.1016/j.procs.2020.06.110.
- [27] Zhu, W., Lan, C., Xing, J., Zeng, W., Li, Y., Shen, L. and Xie, X. 2016. Co-Occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*. i (2016), 3697–3703.