

Klasifikasi Ketertarikan Belajar Anak PAUD Melalui Video Ekspresi Wajah Dan Gestur Menggunakan Convolutional Neural Network

Ajeng Restu Kusumastuti^{[1]*}, Yosi Kristian^[2], Endang Setyati^[3]

Fakultas Sains dan Teknologi^{[1], [2], [3]}

Institut Sains dan Teknologi Terpadu Surabaya

Surabaya, Indonesia

restoeadjenk@gmail.com^[1], yosi@stts.edu^[2], endang@stts.edu^[3]

Abstract—The Covid-19 pandemic has transformed the offline education system into online. Therefore, in order to maximize the learning process, teachers were forced to adapt by having presentations that attract student's attention, including kindergarten teachers. This is a major problem considering the attention rate of children at early age is very diverse combined with their limited communication skill. Thus, there is a need to identify and classify student's learning interest through facial expressions and gestures during the online session. Through this research, student's learning interest were classified into several classes, validated by the teacher. There are three classes: Interested, Moderately Interested, and Not Interested. Trials to get the classification of student's learning interest by teacher validation, carried out by training and testing the cut area of the center of the face (eyes, mouth, face) to get facial expression recognition, supported by the gesture area as gesture recognition. This research has scenarios of four cut areas and two cut areas that were applied to the interest class that utilizes the weight of transfer learning architectures such as VGG16, ResNet50, and Xception. The results of the learning interest classification test obtained a minimum validation percentage of 70%. The result obtained through scenarios of three learning interest classes four cut areas using VGG16 was 75%, while for two cut areas using ResNet50 was 71%. These results proved that the methods of this research can be used to determine the duration and theme of online kindergarten classes.

Keywords— Kindergarten, Facial Expression, Gesture.

Abstrak—Pandemi Covid-19 telah mengubah sistem pendidikan luring menjadi daring. Guru dipaksa beradaptasi dan berinovasi dalam situasi ini. Tuntutan akan tersedianya presentasi yang menarik perhatian siswa, dialami oleh hampir seluruh guru, termasuk guru TK. Materi pembelajaran bagi anak usia dini harus menarik supaya siswa dapat mengikuti proses pembelajaran secara maksimal, sedangkan tingkat ketertarikan anak usia tersebut sangat beragam. Hal tersebut menjadi salah satu kendala yang dihadapi guru TK, selain terbatasnya cara berkomunikasi anak dalam rentang usia tersebut. Oleh karena itu perlu dilakukan klasifikasi ketertarikan belajar melalui ekspresi wajah dan gestur ketika siswa memperhatikan video pembelajaran. Melalui penelitian ini, ketertarikan belajar siswa diklasifikasikan menjadi beberapa kelas yang divalidasi oleh

pengajar. Kelas ketertarikan tersebut dikelompokkan menjadi Tertarik, Cukup Tertarik, dan Tidak Tertarik. Uji coba untuk mendapatkan klasifikasi ketertarikan belajar siswa sesuai dengan validasi pengajar, dilakukan dengan melakukan pelatihan dan pengujian terhadap area potong titik pusat wajah (mata, mulut, wajah) untuk mendapatkan pengenalan ekspresi wajah, didukung area gestur sebagai pengenalan gestur. Dalam menguji hal tersebut, diterapkan skenario 4 area potong dan 2 area potong pada kelas ketertarikan yang memanfaatkan bobot arsitektur transfer learning seperti VGG16, ResNet50, dan Xception. Hasil pengujian klasifikasi ketertarikan belajar memperoleh persentase validasi minimum 70%. Nilai persentase tersebut didapatkan melalui skenario 3 kelas ketertarikan 4 area potong menggunakan VGG16 yaitu sebesar 75%, sedangkan untuk 2 area potong menggunakan ResNet50 dapat mencapai 71%. Hasil ini membuktikan bahwa penelitian dapat digunakan sebagai pertimbangan pengajar dalam menentukan durasi dan tema video bagi siswa TK, selama proses pembelajaran daring.

Kata Kunci—Anak PAUD, Ekspresi Wajah, Gestur.

I. LATAR BELAKANG

Seperti yang disampaikan oleh [1] dalam dinamika Pendidikan anak usia dini, kelompok bermain merupakan representasi nama dari Pendidikan setengah hari bagi anak usia 2-6 tahun. Seperti yang dijelaskan oleh [2] dalam bukunya dengan merujuk pada Pedoman Pendidikan Sekolah oleh Anggini Sudono bahwa pendidikan prasekolah adalah pendidikan untuk membantu pertumbuhan dan perkembangan jasmani dan rohani anak didik di luar lingkungan keluarga sebelum memasuki pendidikan dasar, yang diselenggarakan di jalur pendidikan sekolah atau di jalur pendidikan di luar sekolah. Anak prasekolah merupakan anak dengan rentang usia 4-6 tahun yang merupakan masa peka bagi anak. Anak mulai sensitif untuk menerima berbagai upaya perkembangan seluruh potensi yang dimiliki. Pengembangan kemampuan fisik, kognitif, bahasa, sosial emosional, konsep diri, disiplin, kemandirian, seni, moral, dan nilai agama begitu pesat pada usia ini.

Dalam proses belajar mengajar pada PAUD, dibutuhkan materi yang menarik minat siswa untuk dapat menghasilkan

keluaran yang optimal. Tingkat minat (ketertarikan) atau karakteristik emosi anak dapat ditunjukkan melalui ekspresi wajah [3]. Karena pada dasarnya emosi seseorang adalah perasaan berupa pergolakan pikiran, nafsu, keadaan mental dan fisik yang termanifestasi dalam bentuk gejala takut, cemas, marah dan sebagainya. Pendapat serupa juga dipenjelasan [4], bahwa emosi yang dirasakan oleh anak dapat diekspresikan melalui wajah, dan ekspresi tersebut menyimpan banyak informasi.

Terkait hal tersebut, salah satu dampak dari pandemi COVID-19 adalah transformasi proses pembelajaran dari luring (*offline*) menjadi daring (*online*), sehingga membutuhkan inovasi materi pembelajaran yang menarik. Kebijakan terkait model pembelajaran tersebut telah diterapkan banyak negara, pada seluruh jenjang pendidikan. Mengingat adanya penutupan sekolah sebagai salah satu dampak dari pandemi. [5] Seperti yang diketahui bahwa selama dilakukan pembelajaran daring, beberapa pengajar tertarik untuk memanfaatkan aplikasi *online meeting* dalam proses *transfer knowledge*. Pengajar harus memiliki perilaku efektif dalam menyampaikan pembelajaran yang terpusat pada anak, meskipun tidak dilakukan secara verbal. [6]

Hal tersebut melatarbelakangi peneliti untuk membantu pengajar dalam melakukan evaluasi durasi pembelajaran daring, sesuai dengan keterbatasan interaksi, supaya materi yang disajikan dapat diikuti dengan baik oleh siswa. Penelitian dilakukan dengan mengukur klasifikasi ketertarikan belajar siswa PAUD melalui video ekspresi wajah dan gestur siswa selama video pembelajaran ditayangkan (disajikan melalui monitor).

Partisipan penelitian merupakan anak PAUD TK A berusia 4-5 tahun yang terdaftar di TK IT Harum Jl. Keputih Tegal Bakti III Blok-A Kav No. 110-111. Pemilihan rentang usia partisipan, diuji karena berdasarkan penelitian terkait ekspresi wajah, pada penelitian sebelumnya [7][8][9] berhasil dilakukan pada orang dewasa. Akan tetapi, belum adanya penelitian dengan partisipan anak, didukung dengan ragam tingkat ketertarikan dan keterbatasan cara berkomunikasi yang dimiliki, melatarbelakangi adanya penelitian ini. Proses pelatihan data dilakukan terhadap setiap bagian wajah dan gestur menggunakan *Convolutional Neural Network* dengan memanfaatkan teknik *transfer learning*. Sedangkan pengujian dan validasi dilakukan terhadap mini-video sesuai hasil pengelompokan pengajar.

II. TINJAUAN PUSTAKA

Beberapa penelitian terkait ekspresi wajah telah dilakukan dengan dataset yang melibatkan orang dewasa, diantaranya penelitian [7] yang memanfaatkan *ensemble CNN* dengan 2 *conv layer*, 2 *subsampling layer*, 1 *fully connected layer*, dan 1 *output layer*. Pada keluaran diterapkan 6 *node output* yang didapatkan melalui rata-rata *ensemble* dari *confidence output* beberapa CNN. Iterasi yang digunakan adalah 40 hingga 120 dengan kelipatan 2 disetiap percobaannya. Hasil yang didapatkan menggunakan teknik *ensemble* jauh lebih baik dibandingkan CNN tunggal yaitu 97.4, 97.9, 98, 97.9, dan 98.1

pada setiap iterasi.

Sedangkan penelitian [8] mengungkap MRE-CNN untuk menguji data RAF-DB dengan mengambil 3 hingga 10 frame dengan wajah bersih, menggunakan *adaptive frame interval*. Dalam tahap augmentasi data, dilakukan proses *flipping* dan rotasi setiap 4 hingga 6 derajat, selain menambahkan *gaussian white noise* dengan varian 0.001, 0.01, dan 0.015. Proses pelatihan data menggunakan VGG16 sebanyak 20 iterasi dengan mengatur $lr = 0.00010.0005$, *weight decay* 0.0001, momentum 0.9, *batch size* 16, dan mengoptimalkan *stochastic gradient descent*. Hasil komparasi setiap bagian pada dataset dengan MRE-CNN mencapai persentase 76.73, sedangkan pada AFEW 7.0 dalam menetapkan label ekspresi tunggal dari 7 kelas ekspresi didapatkan persentase sebesar 47.43%.

Pengenalan ekspresi wajah pada penelitian [9] dengan dataset FER2013, diuji dengan 3 subnet yang berbeda. Perbedaan dari setiap subnet tersebut terletak pada lapisan konvolusi. pada tahapan pelatihan, *mini-batch gradient descent* digunakan untuk melatih seluruh subnet dan model. Ukuran *batch* ditetapkan menjadi 100 dengan momentum 0.9, $lr = 0.01$ yang secara bertahap mengalami penurunan hingga 0.001. Iterasi yang digunakan mulai dari 20 hingga 100. Tahapan evaluasi diterapkan *objective function*, top1 akurasi, dan top2 akurasi. Hasil validasi terbaik diperoleh melalui subnet 3 dengan hasil top1 62.44% dan pada top2 memperoleh 79.80%.

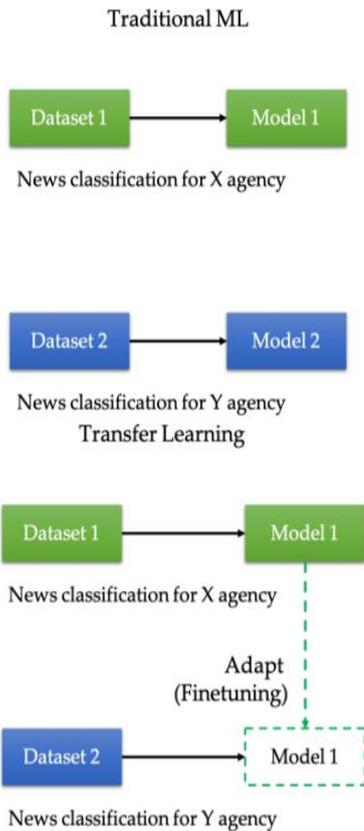
Terkait dengan ketertarikan belajar yang akan peneliti uji terhadap anak PAUD, peneliti [10] melakukan penelitian dengan mempertimbangkan *intake gesture* yang diambil dengan kamera 360 derajat. Deteksi dilakukan dengan menggunakan estimasi probabilitas pada frame level dan deteksi melalui seleksi *sparse poin* waktu berdasarkan perkiraan probabilitas. Model yang digunakan adalah 3D CNN, CNN-LSTM, Two-Stream, dan SlowFast. Pada pendekatan 3D CNN diterapkan ResNet50 *instantiation* yaitu merubah dimensi input dengan menerapkan *stride = 2* pada lapisan konvolusi pertama, selain merubah 2D CNN menjadi 3D CNN. Hasil terbaik dari pengujian seluruh pendekatan, hasil evaluasi F1 pada SlowFast mendapatkan nilai persentase 86%.

Transfer learning adalah suatu teknik yang berisikan model terlatih dalam memecahkan permasalahan lain dengan starting point, memodifikasi dan memperbaharui parameternya sesuai dataset yang baru. [11] Proses pelatihan pada penelitian ini menerapkan bobot arsitektur *transfer learning* seperti VGG16, ResNet50, dan Xception.

VGG merupakan sebuah arsitektur yang terdiri dari blok konvolusi dengan ReLU juga max pooling, dan diakhiri dengan fully connected layer berisikan 1000 kelas menggunakan fungsi softmax. Fungsi tersebut digunakan untuk mendapatkan rentang probabilitas keluaran diantara nilai 0 hingga 1, dan jumlah semua probabilitas akan sama dengan 1. Filter yang digunakan pada arsitektur ini adalah 3x3, karena peneliti [12] mempercayai bahwa lapisan tersebut dapat mencakup area 5x5 dan 7x7. Hal tersebut dibuktikan dengan 1 lapisan filter berukuran 7x7 maka didapatkan parameter sebanyak 49, sedangkan dengan ukuran 3x3 hanya menghasilkan parameter

sebanyak 27, sehingga didapatkan efisiensi hingga 45% dari parameter yang dihasilkan.

Residual Network (ResNet) merupakan arsitektur yang terdiri dari beberapa blok yang dihubungkan oleh *shortcut connection*. [13] Jaringan residual memiliki dua varian yaitu 2 lapisan konvolusi berukuran 3x3 (18-layer dan 34-layer), dan 3 lapisan konvolusi berukuran 1x1, 3x3, dan 1x1 (50-layer hingga 152-layer). Lapisan konvolusi 1x1 yang pertama di varian kedua bertujuan untuk mengurangi dimensi, sedangkan ukuran yang sama setelah 3x3 digunakan untuk mengembalikan dimensi keluaran pada lapisan sebelumnya. Varian kedua ini lebih dikenal dengan istilah *bottleneck* yang bertujuan untuk mengurangi biaya komputasi karena ukuran konvolusi yang lebih kecil.

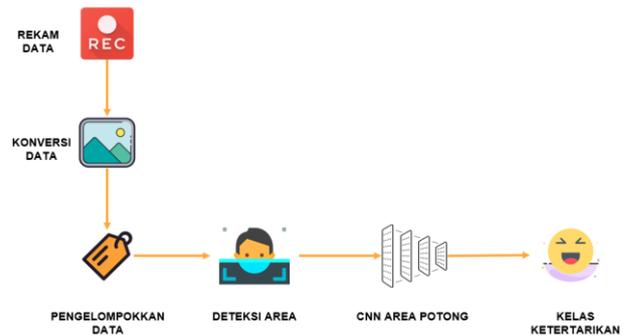


Gambar. 1. Perbandingan Traditional dengan Transfer Learning

Xception atau *Extreme Inception* merupakan tumpukan linear dari depthwise separable lapisan konvolusi dengan residual connection. Sehingga arsitekturnya sangat mudah didefinisikan dan dimodifikasi. Hanya membutuhkan 30 hingga 40 baris kode seperti Keras atau TensorFlow-slim dalam proses modifikasi. Optimalisasi pada depthwise konvolusi, menjadikan Xception memiliki performa lebih cepat dibanding Inception. [14] Arsitektur ini merupakan pengembangan dari Inception yang memiliki 36 lapisan konvolusi untuk membentuk jaringan ekstraksi fitur. Lapisan tersebut merupakan susunan dari 14 modul yang saling terkoneksi

residual linear dari modul 2-13.

III. KLASIFIKASI KETERTARIKAN ANAK MEMANFAATKAN TRANSFER LEARNING



Gambar. 2. Alur Sistem

Proses penelitian klasifikasi anak PAUD berdasarkan ekspresi wajah dan gestur, dilakukan dengan proses awal dan pengujian dalam bentuk video. Bentuk awal data (ekstensi video) yang kemudian dilakukan *sequence of frame* mengadaptasi dari penelitian [9]. Adapun beberapa tahapan yang dilakukan dalam penelitian ini, dijelaskan melalui gambar 2 yang menunjukkan gambaran alur sistem. Langkah pertama yang dilakukan adalah proses pengambilan data dengan melibatkan 35 siswa PAUD TK A berusia 4-5 tahun. Data diambil dengan cara merekam area kepala hingga setengah badan setiap siswa dengan durasi maksimal 5 menit. Dari total siswa tersebut, dilakukan distribusi data sebanyak 80% (28 siswa) yang digunakan sebagai data pelatihan, sedangkan 20% (7 siswa) sisanya digunakan dalam pengujian.

TABEL I. CONTOH KELAS KETERTARIKAN BELAJAR ANAK PAUD

Nama Kelas Ketertarikan Belajar	Contoh Gambar
Tertarik (T)	
Cukup Tertarik (CT)	
Tidak Tertarik (TT)	

Sebelum dilakukan proses pelatihan, seluruh video yang akan dilatih, dikonversi terlebih dahulu ke dalam bentuk gambar, dengan menentukan bahwa setiap 60 detik dapat menghasilkan 1 gambar. Sehingga banyaknya gambar yang dihasilkan dari video setiap siswa sangat beragam sesuai dengan durasi dan proses deteksi. Kemudian, setiap gambar tersebut dilakukan proses pengelompokan data ke dalam tiga kelas ketertarikan yang divalidasi oleh pengajar. Kelas ketertarikan tersebut diantaranya adalah Tertarik, Cukup Tertarik, dan Tidak Tertarik. Pengelompokan data dilakukan dengan contoh kelas hasil validasi pengajar, yang ditunjukkan melalui tabel 1. Proses pengelompokan dibagi menjadi beberapa tahapan yaitu: (1) pengajar memberikan contoh kelas ketertarikan; (2) peneliti mengelompokkan setiap gambar berdasarkan contoh yang disediakan; (3) pengajar memvalidasi setiap kelompok kelas.

A. Pra-Prosesing

Langkah pertama yang dilakukan setelah pengelompokan data adalah deteksi area wajah. Proses deteksi dilakukan dengan memanfaatkan metode Dlib untuk mendapatkan landmark wajah. Sehingga didapatkan potongan area wajah yang digunakan dalam melakukan pengenalan ekspresi seperti area mata, area mulut, dan area wajah. Kemudian dilakukan proses pemotongan area gestur yaitu bagian kepala hingga setengah badan siswa, dengan melakukan pergeseran titik koordinator setelah mendapatkan area wajah. Hasil perpotongan setiap area tersebut dijelaskan melalui gambar 3.



Gambar. 3. Area Potong

Augmentasi data seperti rotasi setiap 15 hingga 20 derajat dan *flipping* dilakukan sebagai upaya penyeimbangan data. Proses *flipping* dilakukan secara horizontal dengan memanfaatkan Image.FLIP_LEFT_RIGHT yang disediakan oleh TensorFlow. Sehingga didapatkan total 12878 gambar dengan rincian sebagai berikut: 5064 untuk kelas Tertarik, 4235 pada kelas Cukup Tertarik dan 3579 pada kelas Tidak Tertarik.

B. Pelatihan Data

Total data yang digunakan dalam pelatihan ini sebanyak 12878 data (sesuai rincian setiap kelas ketertarikan pada poin A) dengan distribusi sebesar 11590 pada data *training*, 644 data validasi, dan 644 data *testing*. Proses distribusi selama pelatihan, dengan memanfaatkan *data split generator*. Pada proses pelatihan data setelah sebelumnya dilakukan proses augmentasi, juga diterapkan objek datagen Data Image

Generator, untuk memudahkan proses pengolahan dataset. Proses pelatihan setiap area potong dilakukan dengan memanfaatkan bobot arsitektur *transfer learning* seperti VGG16, ResNet50, dan Xception. Perubahan arsitektur hanya dilakukan pada *fully connected* layer dengan menyesuaikan kelas menjadi tiga. Data yang diolah pada penelitian ini berukuran 64x64 dengan range warna Red, Green, dan Blue (RGB). Adapun nilai akurasi dan *loss* setiap area potong ditunjukkan melalui tabel II.

TABEL II. NILAI VALIDASI AKURASI DAN LOSS

Model	Mata (%)		Mulut (%)		Wajah (%)		Gestur (%)	
	Acc.	Los.	Acc.	Los.	Acc.	Los.	Acc.	Los.
VGG16	86	36	91	27	97	11	94	20
ResNet50	96	16	87	59	97	11	91	46
Xception	97	28	89	48	97	25	93	35

Proses pelatihan untuk setiap area potong pada setiap arsitektur dilakukan dengan iterasi yang berbeda, yaitu antara 30 hingga 50. Perbedaan iterasi tersebut dilakukan dengan mempertimbangkan hasil peningkatan nilai validasi akurasi dan validasi *loss* di setiap iterasi. Pelatihan juga dilakukan untuk *ensemble CNN*, yaitu tabel yang berisikan kemungkinan nilai ketertarikan dari banyaknya nilai N pada sebuah kelas. Metode ini digunakan dalam pengujian hasil dari pelatihan seluruh area potong, dengan mempertimbangkan bahwa *ensemble* lebih baik dibandingkan *single CNN* seperti yang diketahui dari penelitian sebelumnya [7].

TABEL III. CONTOH TABEL ENSEMBLE CNN

e1	e2	e3	f1	f2	f3	m1	m2	m3	1	2
1	0	0	1	0	0	1	0	0	1	0
0	1	0	1	0	0	1	0	0	1	0
0	0	1	1	0	0	1	0	0	1	0
1	0	0	0	1	0	1	0	0	1	0
1	0	0	0	0	1	1	0	0	1	0
1	0	0	1	0	0	0	1	0	1	0

Seperti contohnya, jika pada suatu gambar untuk seluruh area potong dideteksi sebagai kelas Tertarik, maka nantinya hasil uji akan dinyatakan dengan kelas yang sama. Atau setidaknya jika dari total 4 area, 75%-nya dideteksi sebagai Tertarik maka keseluruhan akan dikelompokkan ke dalam kelas tersebut. Ketentuan tersebut dicontohkan melalui tabel III, dengan “e” merupakan singkatan dari mata, “f” merupakan wajah, dan “g” merupakan gestur. Sedangkan kelas Tertarik dijelaskan dengan angka 1, dan Tidak Tertarik dengan 2. Kemudian, nilai ketentuan 1 merupakan “iya” dan 0 merupakan “tidak”. Salah satu baris ketentuan yang didapatkan untuk kelas

Tertarik pada 3 area potong, sebanyak 7 baris. Hal tersebut didapatkan dengan posisi minimal 2 area selalu memiliki nilai yang sama.

IV. UJI COBA PENELITIAN

Setiap model h5 yang merupakan hasil dari pelatihan menggunakan bobot arsitektur VGG16, ResNet50, dan Xception, diuji terhadap video pengujian. Video pengujian yang digunakan pada penelitian ini, sebanyak 28 mini-video dengan durasi maksimal 60 detik. Seluruh video tersebut berasal dari 7 siswa yang didistribusikan sebagai *testing set*. Langkah pertama adalah proses validasi (*ground truth*) yang dilakukan oleh pengajar terhadap setiap video uji, sehingga didapatkan kelas ketertarikan untuk setiap video tersebut, yang dijelaskan melalui tabel IV.

TABEL IV. GROUND TRUTH

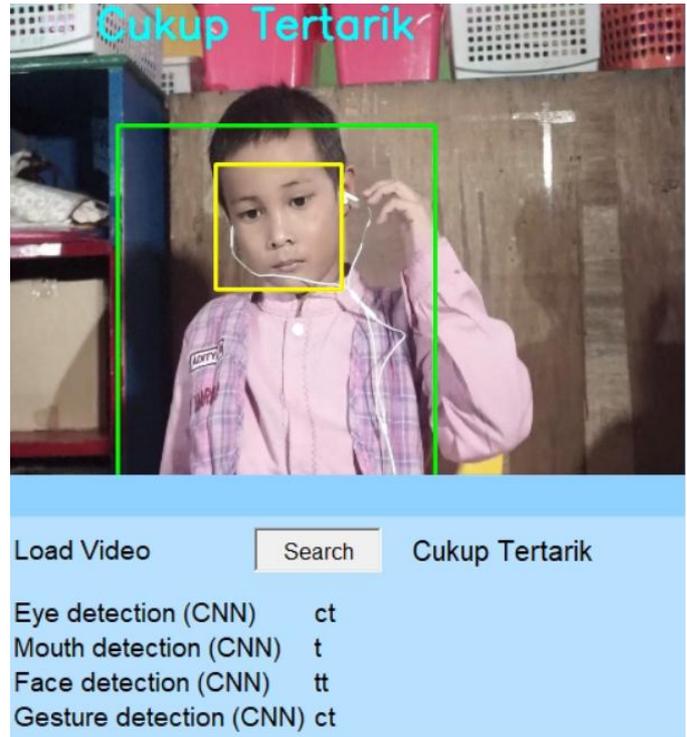
Nama Video	Kelas Ketertarikan Belajar
Video 4, Video 5, Video 6, Video 7, Video 8, Video 15, Video 20, Video 21	Tertarik
Video 1, Video 3, Video 12, Video 13, Video 16, Video 17, Video 22, Video 23, Video 24, Video 25, Video 26, Video 27, Video 28	Cukup Tertarik
Video 2, Video 9, Video 10, Video 11, Video 14, Video 18, Video 19	Tidak Tertarik

Melalui tabel IV didapatkan total 8 video yang divalidasi sebagai kelas Tertarik, 13 video pada kelas Cukup Tertarik, dan sisanya yaitu 7 video pada kelas Tidak Tertarik. Tabel validasi ini pada bab berikutnya akan dijadikan acuan perhitungan validasi *performance metrics* seperti *precision*, *recall*, dan *F1 score* dari setiap hasil pengujian video dengan setiap model *transfer learning*. Contoh tampilan pengujian klasifikasi ketertarikan belajar (skenario 4 area potong) dijelaskan melalui gambar 3.

TABEL V. REKAPITULASI PENGUJIAN 4 AREA POTONG

Kelas	Mata	Mulut	Wajah	Gestur	E
T	0	8	135	2	9
CT	42	96	0	112	99
TT	94	32	1	22	28
Total	136	136	136	136	136

Pengujian yang ditunjukkan melalui gambar 4, merupakan mini-video uji 1 yang divalidasi pengajar sebagai kelas Cukup Tertarik. Berdasarkan contoh pengujian tersebut dengan menerapkan ketentuan tabel *ensemble* seperti yang dicontohkan pada tabel III, maka didapatkan hasil prediksi dari deteksi setiap area potong sebanyak 136 baris. Dari total baris tersebut, didapatkan nilai tertinggi untuk setiap area potong terletak pada kelas Cukup Tertarik, yang dijelaskan melalui tabel V dengan hasil tertinggi pada kolom E sesuai tabel *ensemble*.



Gambar. 4. User Interface

Penamaan kelas ketertarikan pada tabel V menggunakan inisial seperti T merupakan singkatan dari kelas Tertarik, CT adalah kelas Cukup Tertarik, dan TT merupakan Tidak Tertarik. Berdasarkan contoh rekapitulasi pengujian yang dijelaskan pada tabel V, didapatkan hasil validasi *precision*, *recall*, dan *F1-score* untuk total 28 mini-video uji yang dijelaskan melalui tabel VI.

TABEL VI. VALIDASI PENGUJIAN 4 AREA POTONG

Model	Prec.	Rec.	F1
VGG16	0.7505	0.7555	0.7529
ResNet50	0.3274	0.1666	0.2208
Xception	0.6612	0.4305	0.5215

Validasi *F1-score* terbaik yang ditunjukkan pada tabel VI didapatkan melalui VGG16, dengan persentase sebesar 75%. Perhitungan validasi kembali dilakukan untuk setiap area potong seperti mata, mulut, wajah, dan gestur. Hal tersebut dilakukan untuk mendapatkan kombinasi model setiap area potong yang akan diuji kembali sebagai perbandingan hasil akhir dengan pengujian sebelumnya. Hasil validasi setiap area potong disampaikan melalui tabel VII. Melalui tabel tersebut didapatkan kombinasi pengujian model VGG16 untuk gestur dan model Xception untuk ketiga area potong lainnya (mata, mulut, dan wajah). Nilai validasi uji kombinasi adalah 0.65 pada *precision*, 0.68 pada *recall*, dan 0.66 pada *F1-score*.

Terdapat selisih nilai 9% lebih rendah dibandingkan pengujian seluruh area potong menggunakan VGG16.

TABEL VII. VALIDASI PENGUJIAN KOMBINASI 4 AREA POTONG

Area	VGG16(%)			ResNet50(%)			Xception(%)		
	Pre	Re	F1	Pre	Re	F1	Pre	Re	F1
Mata	35	25	29	25	8	12	57	52	54
Mulut	33	10	15	24	46	31	39	31	34
Wajah	54	38	44	54	43	48	51	48	49
Gestur	56	66	61	46	36	41	33	8	13

Sedangkan pada skenario 2 area potong, yang diuji adalah area wajah dan gestur. Kedua area dipilih dengan melihat area terbesar dari pengujian mata manusia, dan karena wajah dianggap telah memberikan informasi yang cukup baik dalam menunjukkan ekspresi partisipan. Rekapitulasi pengujian dengan mini-video yang sama didapatkan hasil pengujian pada kelas Tidak Tertarik. Hal tersebut menunjukkan prediksi yang salah dari tabel IV dengan validasi kelas Cukup Tertarik. Rekapitulasi pengujian untuk skenario ini dijelaskan melalui tabel VIII.

TABEL VIII. REKAPITULASI PENGUJIAN 2 AREA POTONG

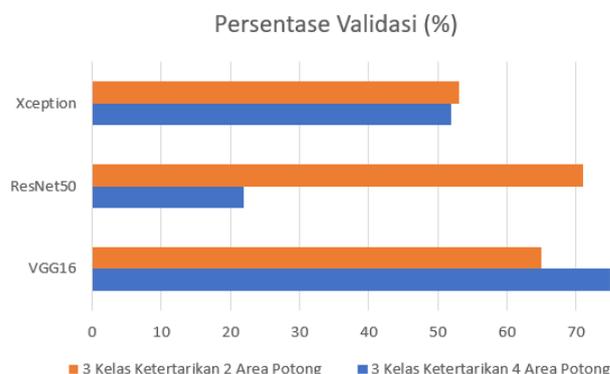
Kelas	Wajah	Gestur	E
T	30	1	30
CT	0	0	0
TT	106	135	106
Total	136	136	136

Berdasarkan hasil validasi menggunakan *F1-score* untuk skenario pengujian 2 area potong, didapatkan nilai tertinggi pada ResNet50. Hasil persentase yang didapatkan selisih 4% lebih rendah, yaitu 71%. Perolehan model VGG16 dan Xception tidak lebih dari 65%. Kemudian dilakukan uji validasi *precision*, *recall*, dan *F1-score* kembali untuk area wajah dan gestur demi mendapatkan kombinasi pengujian yang baru. Uji validasi tersebut (area wajah dan gestur) dijelaskan melalui tabel IX. Tetapi nilai validasi tertinggi diperoleh melalui model VGG16, sehingga pengujian dihentikan karena hasil validasi model tersebut tidak lebih baik dibandingkan ResNet50.

Berdasarkan persentase hasil pengujian menggunakan kedua skenario tersebut (4 area potong dan 2 area potong), didapatkan hasil validasi terbaik pada skenario 3 kelas ketertarikan dengan melibatkan 4 area potong (mata, mulut, wajah, dan gestur). Hal tersebut dijelaskan melalui grafik pada gambar 5.

TABEL IX. VALIDASI PENGUJIAN KOMBINASI 2 AREA POTONG

Area	VGG16(%)			ResNet50(%)			Xception(%)		
	Pre	Re	F1	Pre	Re	F1	Pre	Re	F1
Wajah	63	45	52	54	38	45	51	48	49
Gestur	58	72	64	47	36	41	33	8	13



Gambar 5. Persentase Validasi

KESIMPULAN DAN SARAN

Pada penelitian klasifikasi ketertarikan anak PAUD menggunakan CNN dengan memanfaatkan bobot arsitektur *transfer learning* seperti VGG16, ResNet50, dan Xception diperoleh hasil terbaik pada VGG16. Hal tersebut telah terlihat sejak proses pelatihan, dengan hasil validasi akurasi area mulut dan gestur lebih baik dibandingkan kedua model lainnya. Sedangkan pada area wajah, mendapatkan hasil validasi akurasi yang seimbang, dan validasi akurasi 86% pada area mata. Hal tersebut dibuktikan pada skenario 4 area potong.

Adapun saran untuk penelitian selanjutnya, dapat menambah data penelitian dengan melakukan beberapa sesi pengambilan data, atau dengan melibatkan rentang usia anak yang lebih lebar (usia playgroup yaitu berada di bawah rentang yang diusung peneliti atau TK B dengan rentang 5-6 tahun), ataupun mengarah pada pengujian secara *realtime*.

REFERENSI

- [1] Hary Widodo, "Dinamika Pendidikan Anak Usia Dini", (Semarang: ALPRIN, 2019), hlm.12.
- [2] Dr(C). Irjus Indrawan, Hadion Wijoyo, "Pendidikan Anak Pra Sekolah," (Banyumas: CV. Pena Persada, 2020), hlm. 3.
- [3] Susianty S, Amelia V, Khusniyati M, "Metode Perkembangan Sosial Emosi Anak Usia Dini," Edu Publisher, Tasikmalaya, 2018, cet. 1.
- [4] K Amda, R Fitriyani, "Membaca Ekspresi Wajah," Huta Publisher, Depok, 2016, hlm.17.
- [5] X Bao, H Qu, R Zhang, T P. Hogan, "Modeling Reading Ability Gain in Kindergarten Children during COVID-19 School Closures", IJERPH, vol. 17, 2020.
- [6] M. Agustin, R Dwi Puspita, D. Nurinten, H. Nafiqoh, "Tipikal Kendala Guru PAUD dalam Mengajar pada Masa Pandemi Covid 19 dan Implikasinya". Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini, vol. 5, 2021.

- [7] R. Cui, M. Liu, "Facial Expression Recognition Based on Ensemble of Multiple CNNs". Springer International Publishing AG 2016, pp. 511–518.
- [8] Y. F. B, J. C. K. Lam, and V. O. K. Li, "Multi-region Ensemble Convolutional Neural Network for Facial Expression Recognition". Springer International Publishing, 2018.
- [9] K. Liu, M. Zhang, Z. Pan, "Expression Recognition with CNN Ensemble". International Conference on Cyberworlds, 2016.
- [10] P. V. Rouast, and M. T. P. Adam, "Learning Deep Representastions for Video-Based Intake Gesture Detection," IEEE Journal of Biomedical and Health Informatics, vol. 24, pp. 1727-1737, 2020.
- [11] Samuel Sena, 2018. Pengenalan Deep Learning Part 8: Gender Classification using Pre-Trained Network (Transfer Learning) [Online] <https://medium.com>. [Diakses 24 Agustus 2021].
- [12] Tsang S., 2018. Review: VGGNet – 1st Runner-Up (Image Classification), Winner (Localization) in ILSVRC 2014 [Online] <https://medium.com>. [Diakses 26 Agustus 2021].
- [13] T. Muhammed., "Automated Classification of Histopathology Image Using Transfer Learning", Elseivier: Artificial Intelligence in Medicine, 2019.
- [14] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," IEEE CVPR, vol. 2017, pp. 1800-1807, 2017.