

# Smartphone Photos Categorization Using Markov Model with Limited Training Data

Zulkarnaen Hatala<sup>1</sup>, Muhammad Hudzaly<sup>2</sup>

<sup>1</sup> Program Studi Teknik Informatika, Jurusan Teknik Elektro, Politeknik Negeri Ambon, Ambon, 97234, Indonesia

<sup>2</sup> Jurusan Teknik Industri, Fakultas Teknik Industri, Institut Teknologi Kalimantan, Balikpapan, 76127, Indonesia

## Informasi Artikel

Diterima : 09 Mei 2025  
Revisi : 22 Mei 2025  
Publikasi : 20 Juni 2025

## Kata Kunci:

Smartphone  
Hidden Markov Model (HMM)  
Discrete Cosine Transform (DCT)  
Categorization

## ABSTRAK

Masifnya penggunaan *smartphone* untuk pengambilan foto, maka perlu dilakukan kategorisasi otomatis terhadap foto-foto dalam telepon pintar tersebut. Dalam tulisan ini dipaparkan teknik kategorisasi atau klasifikasi foto dalam *smartphone*. Foto diklasifikasikan menggunakan *Hidden Markov Model* (HMM) dan juga *Siamese Network* untuk melihat akurasi dari kedua model tersebut. Data yang sama digunakan untuk pelatihan dan pengujian untuk kedua model. Untuk HMM kami menggunakan *Discrete Cosine Transform* (DCT) untuk mengekstrak fitur-fitur menonjol dari gambar. Jumlah data sampel pelatihan sangat kecil dibandingkan dengan data sampel pengujian. Di sini kami melakukan metode klasifikasi *few-shot*. Untuk pengenalan HMM, kami menggunakan algoritma Viterbi. Kinerja kedua model diukur untuk hanya 109 sampel uji di luar data pelatihan. HMM mencapai akurasi 98%, Sedangkan *twin network* mencapai 90% dengan sampel pelatih dan pengujian yang sama. Penggunaan HMM memiliki keunggulan dibandingkan Siamese dalam hal komputasi yang lebih cepat. HMM menungknikan telepon pintar dengan kemampuan komputasi rendah untuk mengkategorikan foto secara otomatis. Kami menyediakan kode program tulisan ini di <https://github.com/dzhatala/hmmphoto>.

## ABSTRACT

This writing investigates the classification of images taken using a smartphone. Due to the large number of photos and the large number of photo categories, it is necessary to automatically categorize these photos. Photos are classified using two different approaches. The first method uses Hidden Markov Model (HMM) and the second technique employs Siamese Network from Convolutional Neural Network (CNN) architecture. The same data are used for training and testing for both models. For HMM we use Discrete Cosine Transform (DCT) to extract salient features of images. The number of training examples is very small compared to the test set. Here we carried out few-shot classification method. For recognition of the HMM, Viterbi algorithm is applied. Performances of both procedures were measured. For only 109 test samples HMM achieve 98% accuracy, while twin network achieves 90%. The use of HMM has advantage over Siamese in term of faster computation. HMM opens the opportunity of the smartphone with low computation capability to categorize photos automatically. We provide source code at <https://github.com/dzhatala/hmmphoto>.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



## \*Penulis Koresponden

Email: [dzulqarnaenhatala@gmail.com](mailto:dzulqarnaenhatala@gmail.com)   

Cara sitasi ::

---

Z. Hatala, M. Hudzaly, "Smartphone photos categorization using Markov model with limited limited training data," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 2, p. 573-579, Juni 2025. doi:10.30811/jaise.v5i2.6943

---

## 1. PENDAHULUAN

Saat ini, berbagai fenomena suatu kejadian dapat dengan mudah diabadikan menggunakan telepon seluler. Kamera di dalam telepon pintar mampu mengambil foto dengan lebih baik dan lebih mudah di mana saja dan kapan saja tanpa biaya. Mengambil gambar dengan kamera telah menjadi kebiasaan orang di seluruh dunia. Satu orang secara statistik dapat memiliki rata-rata 640 foto yang diambil oleh telepon pintarnya [1]. Selain fungsinya sebagai kamera, telepon pintar mengintegrasikan kemampuannya untuk memproses lebih lanjut dan mengolah dokumen gambar. Seorang pengguna telepon pintar dengan profesi sebagai pendidik seperti profesor, dosen, atau guru dapat mengakses secara digital tugas yang diserahkan siswa menggunakan telepon pintar. Keuntungan menggunakan telepon daripada pemindai (*scanner*) *flatbed* khusus adalah mobilitas dan fleksibilitas. Dengan telepon berbasis kamera yang mudah digunakan, yang menghasilkan begitu banyak gambar yang diambil untuk setiap orang, tentunya ada kebutuhan untuk mengelola dan mengatur gambar-gambar ini secara efisien untuk banyak tujuan yang diperluas. Seorang profesor perlu menfoto semua kertas ujian mahasiswa dan membedakannya dari jenis foto lain seperti panorama, acara keluarga, upacara, dll. Karena gambar dalam telepon pintar sangat banyak, meskipun klasifikasi gambar serupa ke dalam kategorinya cukup mudah, proses pengulangan bisa jadi sangat melelahkan untuk dilakukan oleh manusia. Di sisi lain, telepon pintar dan komputer dapat diprogram untuk melakukan klasifikasi secara otomatis. Keduanya lebih cepat dan tidak membuat lelah. Selain itu, orang tersebut bebas memindahkan gambarnya ke perangkat pemrosesan yang lebih baik seperti laptop berperforma tinggi untuk memperoleh waktu eksekusi yang lebih singkat dan akurasi yang lebih tinggi. Kegiatan mengkategorikan gambar menggunakan teknik dalam ilmu *machine learning*. Gambar yang diambil oleh telepon pintar ini agak sulit dikenali daripada gambar yang diperoleh menggunakan pemindai *flatbed* khusus. Ini karena foto telepon pintar terdistorsi oleh gerakan tangan manusia variasi intensitas pencahayaan.

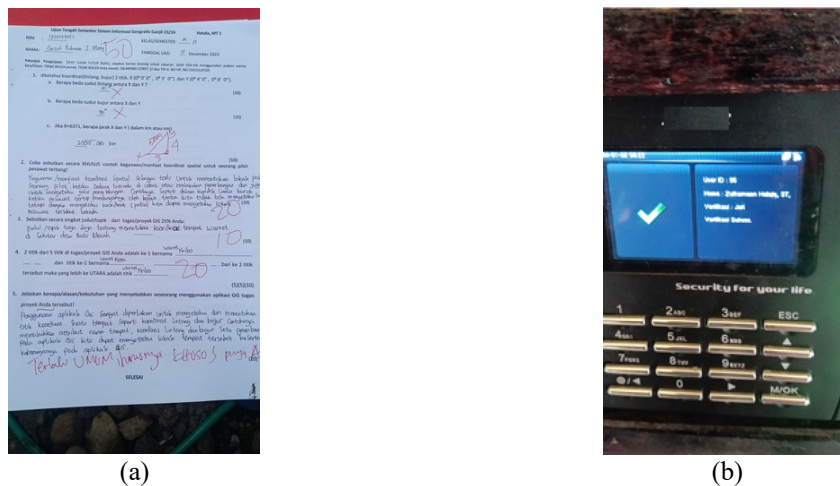
Hidden Markov Model (HMM) [1], [2] adalah metode statistik yang telah digunakan dalam banyak aplikasi dunia nyata. Secara historis, HMM telah memperoleh keberhasilan dalam bidang topik yang berhubungan dengan ucapan, termasuk pengenalan[2], dan verifikasi[3]. Akhir-akhir ini, HMM telah diterapkan dalam banyak penelitian klasifikasi gambar. HMM yang diterapkan pada pengenalan wajah dengan jumlah komputasi yang mungkin dibahas di [4]. HMM bersama dengan *Neural Network* (NN), *Principle Component Analysis* (PCA), dan *Gabor Filter* (GF) dibahas dalam [5]. HMM juga digunakan untuk mengkategorikan gambar pakaian [6]. Klasifikasi gambar *thumbnail* resolusi rendah kecil dilakukan di [7]. Di sisi lain, *Neural Network* juga digunakan secara luas dalam klasifikasi gambar. Klasifikasi dengan sampel yang sangat kecil dibandingkan dengan sampel uji dilakukan dengan banyak model dari berbagai arsitektur. Residual Network (ResNet) digunakan dalam [8]. Adversarial network digunakan dalam[9]. Transformer digunakan dalam [10], [11], [12]. Arsitektur *Siamese twin network* digunakan untuk mengklasifikasikan huruf-huruf tulisan tangan dilaporkan di [13].

Dalam penelitian ini, gambar yang diambil menggunakan telepon pintar dikategorikan menggunakan dua pendekatan berbeda yang disebutkan di atas, Model Markov Tersembunyi dan Jaringan Syaraf Tiruan. Saat menggunakan HMM, kami menggunakan *Discrete Cosine Transform* (DCT) [14] sebagai teknik ekstraksi fitur. Sejumlah kecil koefisien DCT harus menangkap fitur menonjol sederhana yang meniru proses pengenalan manusia atau bahkan pola lebih halus yang terlewatkan oleh mata manusia. Kesamaan lokal dari baris dan kolom yang berdekatan menunjukkan ketergantungan blok antar-tetangga di dalam suatu gambar. Fakta-fakta ini memotivasi klasifikasi foto menggunakan HMM. Pendekatan lain yang kami gunakan adalah Siamese Network [13] a type of *Convolutional Neural Network* (CNN) yang juga diterapkan dalam penelitian ini sebagai perbandingan dengan metoda HMM. Dalam makalah ini, gambar yang diambil menggunakan telepon pintar dikategorikan menggunakan dua pendekatan berbeda yang disebutkan di atas, Model Markov Tersembunyi (HMM) dan Jaringan Syaraf Tiruan (*Artificial Neural Network*). Saat menggunakan HMM, kami menggunakan DCT [14] sebagai teknik ekstraksi fitur. Sejumlah kecil koefisien DCT harus menangkap fitur menonjol sederhana yang meniru proses pengenalan manusia atau bahkan untuk pola yang lebih halus yang terlewatkan oleh mata manusia. Kesamaan lokal dari baris dan kolom yang berdekatan menunjukkan ketergantungan blok antar-tetangga di dalam suatu gambar. Fakta-fakta ini memotivasi klasifikasi foto menggunakan HMM.

## 2. METODE

### 2.1 Data

Dalam penelitian ini semua gambar diambil menggunakan kamera *smartphone*, menggunakan aplikasi media sosial Whatsapp. Pemilik telepon tersebut adalah seorang dosen. Kami membatasi jenis gambar hanya menjadi dua kategori yaitu gambar mesin sidik jari (Gambar 1.a) dan hasil *scan* kertas ujian mahasiswa (Gambar 1.b). Sebagai seorang karyawan dosen harus mencatat kehadirannya di mesin biometrik *fingerprint* setiap hari kerja di universitas. Karena universitas belum menerbitkan konfirmasi rekapitulasi kehadiran, sang dosen mengambil foto sidik jari untuk bukti pribadi. Dalam kasus kedua, seorang pendidik secara teratur menjalankan tes berbasis kertas untuk sesi ujian di kelas binaannya. Setiap kali mereka selesai menilai makalah, mereka mengambil foto kertas ujian dan menyimpannya di *smartphone*. Hal ini berguna agar versi digital dari kertas tersebut dapat dikirim kembali ke mahasiswa. Contoh gambar-gambar ini ditunjukkan pada Gambar 1.



Gambar 1. Contoh data pelatihan (a) Tulisan tangan ujian mahasiswa dan (b) layar mesin fingerprint.

Dari semua foto, maka beberapa dipilih secara acak kemudian diberi *label* atau anotasi. Data yang beranotasi ini berfungsi sebagai data pelatihan yang akan digunakan oleh kedua model, HMM dan *twinn network*. Pemisahan antara data pengujian dan pelatihan bergantung pada ketersediaan label (anotasi) suatu foto. Hanya data yang memiliki label yang dapat dimasukkan ke dalam sampel pelatihan.

### 2.2 Metoda klasifikasi dengan HMM

Bagian ini menjelaskan metode untuk mengklasifikasikan gambar yang ada menggunakan HMM. Tahapannya dirangkum di bawah ini:

- a. Konversi warna gambar: semua gambar diubah menjadi gambar tingkat abu-abu 8 bit.
- b. Normalisasi ukuran gambar: foto diubah ukurannya (diperkecil) secara proporsional ke lebar tetap dan panjang variabel. Sebagian besar gambar yang diambil oleh telepon pintar jauh lebih besar daripada ukuran yang diperkecil. Penskalaan ukuran gambar dimaksudkan untuk mencapai waktu pemrosesan yang lebih cepat dan agar sesuai dengan model. Dalam penelitian ini semua gambar diubah ukurannya menjadi lebar konstan 280 piksel. Tinggi gambar bervariasi tergantung pada ukuran asli foto untuk menjaga agar foto tetap proporsional.
- c. Ekstraksi fitur: gambar dibagi menjadi subblok dengan ukuran blok yang tetap. Seperti pada Gambar 2, ditunjukkan subblok yang tidak tumpang tindih (*non overlap*). Blok yang tumpang tindih saat ini belum digunakan. Fitur diekstraksi untuk setiap blok dan fitur ini diasumsikan diekstraksi oleh *state* HMM. Koefisien *Discrete Cosine Transform* (DCT) dihitung untuk setiap blok [15] [16] [17]. Jumlah blok untuk sebuah gambar bervariasi tergantung pada panjang gambar. Seperti yang ditunjukkan pada gambar 1. sepanjang arah horizontal dari kiri ke kanan kita memiliki C jumlah blok untuk satu baris. Pada arah vertikal kita memiliki total R blok untuk kolom dari  $b_{00}$  hingga  $b_{R-1,C-1}$ . Jadi jumlah total fitur DCT untuk sebuah gambar adalah  $R \times C$ . Selain itu, di setiap ujung baris, fitur khusus ditambahkan yang disebut MARKER [18] diposisikan pada blok imajiner M pada gambar 2.

$b_{00}$	$b_{01}$				$b_{0c-1}$	M
$b_{10}$	$b_{11}$				$B_{1c-1}$	M
			$b_{rc}$			
$b_{R-10}$	$b_{R-11}$				$b_{R-1c-1}$	M

Gambar 2. Pembagian subblok untuk ekstraksi fitur pada foto

Pada penelitian ini ukuran blok yang dipilih adalah 4x4 pixel tanpa *overlap*. Koefisien DCT pada setiap block dihitung menurut [15]:

$$F(u, v) = \frac{C(u)C(v)}{2} \sum_0^3 \sum_0^3 p(ij) \cos\left(\frac{(2i+1)u\pi}{8}\right) \cos\left(\frac{(2j+1)v\pi}{8}\right) \quad (1)$$

dengan

$$C(m) = \begin{cases} \frac{\sqrt{2}}{2} & \text{if } m = 0 \\ 1 & \text{otherwise} \end{cases}$$

dan  $p(ij)$  level *grayscale* untuk pixel dengan koordinat  $i, j$  pada blok tersebut. Dalam subblok  $b$  dari kemungkinan 16 koefisien DCT yang dihitung cuma 6 yang digunakan membentuk fitur [17] yaitu pada persamaan (2) berikut:

$$\begin{aligned} f_{b0} &= F(0,0), f_{b1} = F(0,1), f_{b2} = F(1,0), \\ f_{b3} &= F(2,1), f_{b4} = F(2,2), f_{b5} = F(1,2) \end{aligned} \quad (2)$$

Kemudian ditambahkan juga 10 koefisien yaitu koefisien=koefisien Delta dan Akselerasi. Koefisien delta adalah selisih dari koefisien DCT antara blok pada dua kolom yang berdekatan dalam satu baris yang sama. Serupa dengan delta, koefisien-koefisien akselerasi adalah perbedaan antara delta pada dua subblok yang berdekatan kolom dalam baris yang sama. Baik delta maupun akselerasi bisa ditulis dalam persamaan (3) berikut:

$$f_{bc} = f_{b,c-5} - f_{b-1,c-5}, \quad c = 6...15 \quad (3)$$

Delta dan akselerasi pada blok-blok di kolom pertama diinisialisasi dengan nilai nol. Dengan demikian kita memiliki *observation vector* dengan berisikan 16 nilai terdiri dari 6 DCT, 5 delta and 5 koefisien akselerasi. Vektor observasi adalah output yang dihasilkan oleh suatu *hidden state* dari HMM.

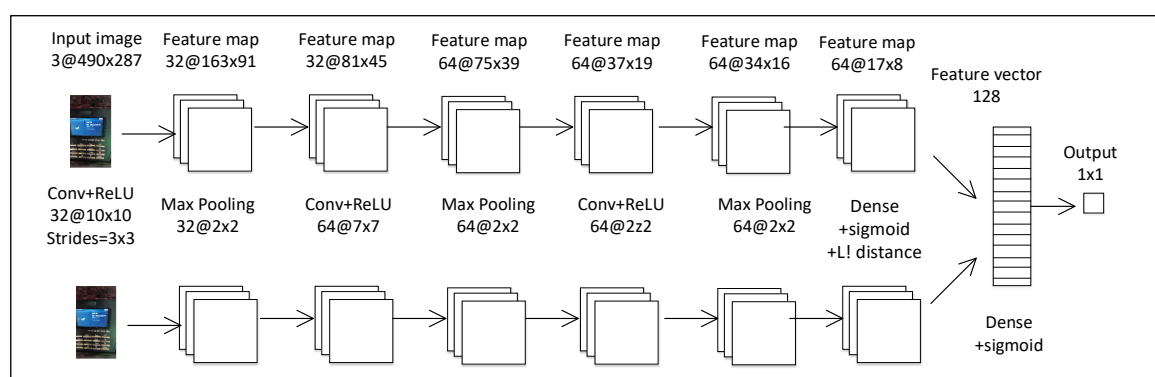
d. Pelatihan Model HMM (*Training*). Kami menggunakan kelas HMM dengan topologi tetap untuk membuat model sepanjang baris (arah horizontal gambar). Masing-masing dengan 4 status eksitasi (status yang memancarkan observasi). HMM ini akan membuat model semua vektor fitur sepanjang baris. Pertama, satu HMM diinisialisasi menggunakan algoritma pelatihan Viterbi [19] [20] [21], yang digunakan pada semua data pelatihan. Inisialisasi dilakukan menggunakan program *HInit*. Kemudian HMM ini disalin ke enam HMM secara berurutan untuk setiap baris. Kemudian *HRest* yang menjalankan algoritma Baum Welch yang terisolasi digunakan untuk melatih HMM individual berdasarkan data. Akhirnya, semua HMM digabungkan menjadi satu HMM besar dan dilatih menggunakan semua foto yang tersedia. *Embedded training* dilakukan menggunakan program dari HTK yaitu *HERest*.

e. Klasifikasi gambar:, algoritma Viterbi menghitung probabilitas gambar terhadap semua kemungkinan kategori. Varian algoritma Viterbi yang digunakan disebut token passing [22] [23]. Jika terdapat beberapa

kategori pengenalan, algoritma pengenalan Viterbi akan menentukan probabilitas setiap gambar terhadap semua jalur kategori yang ada. Kategori gambar uji diputuskan berdasarkan skor viterbi terbaik. Program HTK HVite digunakan untuk menjalankan algoritma klasifikasi.

## 2.2 Metoda klasifikasi dengan *Siamese (Twin) Network*

Untuk membandingkan hasil dengan HMM, kami menggunakan arsitektur *Siamese network* [13] [24] sebagaimana tergambar pada Gambar 3. Jaringan kembar (*twin network*) adalah jenis CNN yang menerapkan teknik *deep learning* [25]. Penggunaan pendekatan beberapa lapisan tersembunyi yang diterapkan di sini terutama dijelaskan dalam [13]. Metode dalam [8] memerlukan dimensi lebar dan tinggi foto yang tetap. Setelah mengubah ukuran secara proporsional dari gambar asli, gambar tersebut secara dipotong atau diberi tambahan di sisi kiri dan kanan agar memenuhi ukuran yang diinginkan arsitektur *twin network*. Hal ini karena dimensi gambar yang diteruskan ke lapisan input harus tetap yaitu dengan ukuran 498x280 piksel. Warna gambar juga diteruskan sebagai dimensi ke-3 dari *tensor* [26]. Informasi warna ini berisi tiga saluran untuk merah, hijau dan biru (*RGB*). Saat gambar diteruskan ke lapisan input, gambar tersebut kemudian diproses melalui lapisan-lapisan (*layers*). *Twin network* berisi lapisan konvolusi (*Convolution layers*) yang melakukan ekstraksi fitur otomatis [27].



Gambar 3. Arsitektur *Siamese Network* pada kategorisasi foto

Pada gambar 3, input ke arsitektur adalah 2 buah foto dikirim masing masing ke layer pertama yang berupa *convolutional layer* dengan *kernel size*=10x10, jumlah filter sebanyak 32 dan *strides*=3x3, dengan fungsi aktivasi ReLU (32@10x10 teks pada gambar bagian tengah). Hasil ini menghasilkan feature map berjumlah 32 dengan ukuran 163x91 (32@163x91 terlihat pada teks gambar paling atas). Setelah itu output dari lapis konvolusi diteruskan ke lapisan *max pooling* menghasilkan feature map sebanyak 32 dengan ukuran 81x45 (32@81x45). Demikian seterusnya hingga pada lapisan terakhir yaitu 128 *feature vector* dilewatkan ke lapisan *dense* dan aktivasi sigmoid untuk menghasilkan 1 nilai similaritas antara 2 input foto *smartphone*.

Kami menggunakan *library tensorflow* dan *keras* [28] untuk mengimplementasikan seluruh model, pelatihan dan klasifikasi. Untuk melatih *Siamese network* dengan hanya dua sampel yang diambil dari dua kategori, kita akan memiliki ukuran *batch* enam dengan jumlah *batch* hanya satu, yaitu 6 sampel dilewatkan sekaligus dalam satu siklus *forward* sebelum siklus *feedback* dijalankan untuk memperbaharui parameter. Setelah mengulang periode ini sekitar 7000 siklus (*epoch*), parameter jaringan yang dihasilkan kemudian digunakan untuk menguji 109 sampel lainnya.

## 3. HASIL DAN PEMBAHASAN

Kami menggunakan 113 foto yang terdiri dari dua kategori, lembar jawaban siswa, dan mesin sidik jari (*fingerprint*). Sebanyak 55 foto termasuk dalam lembar jawaban siswa dan sisanya adalah mesin *fingerprint*. Dari masing-masing dua kelas, kami hanya mengambil dua contoh untuk tujuan pelatihan. Hasil klasifikasi dua bidikan untuk kedua metode disajikan dalam tabel di bawah ini. HMM menggunakan ekstraksi fitur terpisah di luar proses pelatihannya, sedangkan ekstraksi fitur CNN disematkan otomatis dalam arsitektur neural network tersebut. Kami menggunakan 113 foto yang terdiri dari dua kategori, lembar jawaban siswa, dan papan sidik jari. Sebanyak 55 foto termasuk dalam lembar jawaban siswa dan sisanya adalah papan. Dari masing-masing dua kelas, kami hanya mengambil dua contoh untuk tujuan pelatihan. Hasil klasifikasi dua bidikan untuk kedua metode disajikan dalam tabel di bawah ini. HMM menggunakan ekstraksi fitur terpisah di luar proses pelatihannya, sedangkan ekstraksi fitur CNN disematkan dalam proses pelatihan. Dari tabel 1, performansi akurasi HMM dan CNN bisa dikatakan berhasil dengan tingkat akurasi di atas 50%. Dari tabel 1, HMM dan CNN dapat dibandingkan yaitu HMM mempunyai akurasi 98 persen yaitu benar mengenali 107

foto dan salah mengenali 2 foto. Sedangkan CNN lebih kecil akurasi yaitu 90% mampu benar mengenali 97 foto dan salah 12 foto.

Tabel 1. Hasil Akurasi kategorisasi

No	Method	Preprocessing	Feature Extraction	Jumlah kelas	Jumlah sampel pelatih	Jumlah sampel uji	Akurasi
1	Hidden Markov Model	Convert images to black and white	DCT	2	4	109	98%
2	Siamese Network	Left right Padding	Automatic	2	4	109	90%

*Confusion matrixes* dari kedua metoda ditunjukkan pada gambar tabel 2 dan tabel 3. Pada tabel2, dengan HMM, mesin fingerprint salah dikenali sebagai kertas ujian adalah sebanyak 1 sampel. Hal ini lebih tinggi pada *Siamese* yaitu mesin fingerprint tidak dikenali sebanyak 12 sampel sebagaimana terlihat pada tabel 3.

Tabel 2. *Confusion matrix* dari metoda HMM

Benar-- Dikenali	Kertas ujian	Mesin fingerprint
Kertas ujian	54	0
Mesin fingerprint	1	54

Tabel 3. *Confusion matrix* dari metoda Siamese

Benar-- Dikenali	Kertas ujian	Mesin fingerprint
Kertas ujian	55	12
Mesin fingerprint	0	42

#### 4. KESIMPULAN

Klasifikasi gambar pribadi di dalam telepon pintar adalah tugas yang berguna yang dapat dilakukan dengan menggunakan teknik pembelajaran mesin. Penelitian ini menerapkan dua metode klasifikasi gambar berdampingan, menggunakan HMM dan *Deep Learning Convolutional Siamese Twin Network*. Pertama menggunakan HMM, kami menggunakan DCT sebagai teknik ekstraksi fitur. Sementara tahap ekstraksi fitur diperlukan dalam metode HMM, CNN sebaliknya tidak memerlukan langkah ini. Proses tambahan ini membuat CNN mengungguli HMM dalam kompleksitas prosedur. Dalam aspek lain, CNN memerlukan kecepatan pemrosesan yang lebih tinggi dan memori perangkat komputasi yang lebih besar, sementara HMM umumnya lebih cepat. HMM juga dapat menoleransi bobot piksel dan lebar gambar yang sedikit berbeda, sementara Siamese memerlukan dimensi tetap untuk semua data pelatihan dan pengujian. Kami juga mencatat bahwa ketika HMM digunakan, data pelatihan harus memiliki "label" yang seharusnya dilakukan secara manual untuk menandai bagian-bagian gambar. Anotasi dan pelabelan manual tidak diperlukan saat menggunakan Siamese. Dari hasil klasifikasi, kami mencatat bahwa HMM sedikit lebih baik daripada CNN Siamese dalam hal akurasi. Komputasi yang lebih rendah yang digunakan oleh HMM dalam hal prosesor dan memori memungkinkan penerapan aplikasi pada perangkat telepon pintar spesifikasi rendah yang dijual murah. Kedua model ini berhasil mengklasifikasikan foto pada dataset yang digunakan dengan akurasi tinggi di atas 70%.

#### REFERENSI

- [1] B. Mor, S. Garhwal, and A. Kumar, "A Systematic Review of Hidden Markov Models and Their Applications," *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 1429–1448, May 2021, doi: 10.1007/s11831-020-09422-4.
- [2] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, no. 2, 1989.
- [3] I. B. Kadhim, A. N. Nasret, and Z. S. Mahmood, "Enhancement and modification of automatic speaker verification by utilizing hidden Markov model," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 3, p. 1397, Sep. 2022, doi: 10.11591/ijeecs.v27.i3.pp1397-1403.
- [4] D. Ali, I. Touqir, A. M. Siddiqui, J. Malik, and M. Imran, "Face Recognition System Based on Four State Hidden Markov Model," *IEEE Access*, vol. 10, pp. 74436–74448, 2022, doi: 10.1109/ACCESS.2022.3188717.

- [5] A. Aggarwal, M. Alshehri, M. Kumar, P. Sharma, O. Alfarraj, and V. Deep, "Principal component analysis, hidden Markov model, and artificial neural network inspired techniques to recognize faces," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 9, p. e6157, 2021.
- [6] S. Tena, R. Hartanto, and I. Ardiyanto, "Content-based image retrieval for fabric images: A survey," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 3, p. 1861, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1861-1872.
- [7] M. Mouret, C. Solnon, and C. Wolf, "Classification of Images Based on Hidden Markov Models," in *2009 Seventh International Workshop on Content-Based Multimedia Indexing*, Chania, Crete: IEEE, Jun. 2009, pp. 169–174. doi: 10.1109/CBIMI.2009.22.
- [8] C. Wang, Z. Yu, Z. Long, H. Zhao, and Z. Wang, "A few-shot diabetes foot ulcer image classification method based on deep ResNet and transfer learning," *Sci. Rep.*, vol. 14, no. 1, pp. 1–9, 2024.
- [9] J. Dong, Y. Wang, J.-H. Lai, and X. Xie, "Improving adversarially robust few-shot image classification with generalizable representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9025–9034.
- [10] B. M. S. Maia *et al.*, "Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer," 2024.
- [11] K. Dunphy, M. Buwaneswaran, K. Grolinger, and A. Sadhu, "Few-Shot Learning Augmented with Image Transformation for Multiclass Structural Damage Classification," *J. Comput. Civ. Eng.*, vol. 39, no. 3, p. 04025021, 2025.
- [12] W. Song and Y. Huang, "Adaptive feature recalibration transformer for enhancing few-shot image classification," *Vis. Comput.*, pp. 1–15, 2025.
- [13] G. Koch, R. Zemel, R. Salakhutdinov, and others, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, Lille, 2015.
- [14] M. Hamid, "DCT-based image feature extraction and its application in image self-recovery and image watermarking," PhD Thesis, Concordia University, 2016.
- [15] L. Tan and J. Jiang, *Digital signal processing: fundamentals and applications*, 3rd ed. Academic press, 2019.
- [16] V. V. Kohir and U. B. Desai, "Face recognition using a DCT-HMM approach," in *Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201)*, Princeton, NJ, USA: IEEE Comput. Soc, 1998, pp. 226–231. doi: 10.1109/ACV.1998.732884.
- [17] J. Li and R. M. Gray, *Image segmentation and compression using hidden Markov models*, vol. 571. Springer Science & Business Media, 2000.
- [18] F. S. Samaria, "Face recognition using hidden Markov models," PhD Thesis, University of Cambridge Cambridge, UK, 1994.
- [19] L. R. Rabiner, J. G. Wilpon, and B.-H. Juang, "A segmental k-means training procedure for connected word recognition," *ATT Tech. J.*, vol. 65, no. 3, pp. 21–31, 1986.
- [20] G. D. Forney, "The viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [21] A. Allahverdyan and A. Galstyan, "Comparative analysis of viterbi training and maximum likelihood estimation for hmms," *Adv. Neural Inf. Process. Syst.*, vol. 24, 2011.
- [22] S. J. Young, N. H. Russell, and J. H. S. Thornton, "Token Passing: a Simple Conceptual Model for Connected Speech Recognition Systems".
- [23] R. Solera-Ureña, J. Padrell-Sendra, D. Martín-Iglesias, A. Gallardo-Antolín, C. Peláez-Moreno, and F. Díaz-de-María, "SVMs for Automatic Speech Recognition: A Survey," in *Progress in Nonlinear Speech Processing*, vol. 4391, Y. Stylianou, M. Faundez-Zanuy, and A. Esposito, Eds., in Lecture Notes in Computer Science, vol. 4391, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 190–216. doi: 10.1007/978-3-540-71505-4\_11.
- [24] H. Shao, D. Zhong, X. Du, S. Du, and R. N. Veldhuis, "Few-shot learning for palmprint recognition via meta-siamese network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.
- [25] Y. Lai, "A comparison of traditional machine learning and deep learning in image recognition," in *Journal of Physics: Conference Series*, IOP Publishing, 2019, p. 012148.
- [26] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, 2017.
- [27] T. Wiatowski and H. Bölskei, "A mathematical theory of deep convolutional neural networks for feature extraction," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1845–1866, 2017.
- [28] O. Dürr, B. Sick, and E. Murina, *Probabilistic deep learning: With python, keras and tensorflow probability*. Manning Publications, 2020.