



Implementasi Optical Mark Recognition dengan Pendekatan Template-Based untuk Koreksi Otomatis Lembar Jawaban Pilihan Ganda

Implementation of Optical Mark Recognition Using a Template-Based Approach for the Automated Grading of Multiple-Choice Answer Sheets

Novi Tristanti^{1,*}, Oki Derajat Sudarmojo², Anang Susilo², Anny Muthmainah¹

¹ Fakultas Sains dan Teknologi, Program Studi Informatika, Universitas Muhammadiyah Karanganyar, Karanganyar, Indonesia

² Fakultas Informatika dan Kesehatan, Program Studi Informatika, Universitas Sragen, Sragen, Indonesia

Email: ^{1,*}novitristanti@umuka.ac.id, ²okiderajatsudarmojo@gmail.com, ³anangsusilo.unissra@gmail.com,

⁴annymuthmainah@gmail.com

Email Penulis Korespondensi: novitristanti@umuka.ac.id

Received: 27/0/2026, Revised: 09/10/2026, Accepted: 13/10/2026, Published: 13/10/2026

Abstrak—Pemeriksaan lembar jawaban pilihan ganda secara manual membutuhkan waktu dan tenaga, terutama ketika jumlah lembar jawaban yang diperiksa cukup banyak. Optical Mark Recognition (OMR) dapat digunakan untuk mengotomatisasi proses pembacaan dan koreksi lembar jawaban berbasis citra. Penelitian ini mengimplementasikan sistem OMR berbasis computer vision dengan memanfaatkan template lembar jawaban kosong sebagai acuan dalam mengidentifikasi perubahan tanda pada area pilihan jawaban. Sistem dikembangkan menggunakan bahasa pemrograman Python dan pustaka OpenCV melalui tahapan preprocessing, grayscale, adaptive thresholding, deteksi kontur, koreksi perspektif, penyalarsan citra terhadap template, segmentasi area jawaban, identifikasi jawaban, koreksi berdasarkan kunci jawaban, dan perhitungan nilai. Sistem dirancang untuk mengenali enam kelas keluaran, yaitu A, B, C, D, E, dan KOSONG. Evaluasi dilakukan menggunakan 50 citra lembar jawaban dengan masing-masing 60 soal dan lima alternatif jawaban, sehingga diperoleh 3.000 keputusan jawaban dan 15.000 area pilihan yang dianalisis. Seluruh 50 lembar jawaban berhasil diproses pada tahap validasi. Hasil evaluasi menunjukkan bahwa 2.863 prediksi berhasil dikenali dengan benar dan 137 prediksi salah, dengan accuracy sebesar 95,43%, macro precision sebesar 95,62%, macro recall sebesar 94,53%, dan macro F1-score sebesar 94,92%. Evaluasi khusus kelas KOSONG menghasilkan precision sebesar 100%, recall sebesar 85,71%, dan F1-score sebesar 92,31%. Hasil tersebut menunjukkan bahwa pendekatan OMR berbasis computer vision dan template dapat digunakan untuk melakukan pembacaan dan koreksi otomatis lembar jawaban dengan kinerja yang baik. Meskipun demikian, masih terdapat 18 jawaban yang sebenarnya kosong tetapi teridentifikasi sebagai salah satu pilihan A–E. Kondisi tersebut menunjukkan bahwa variasi visual pada area jawaban masih dapat memengaruhi proses identifikasi, sehingga optimasi mekanisme pembandingan template dan pengujian pada kondisi citra yang lebih beragam diperlukan untuk meningkatkan robustness sistem.

Kata Kunci: Optical Mark Recognition; Computer Vision; Lembar Jawaban Komputer; Pengolahan Citra; Koreksi Otomatis

Abstract—Manual checking of multiple-choice answer sheets is time-consuming and labor-intensive, particularly when a large number of answer sheets need to be examined. Optical Mark Recognition (OMR) can be used to automate the image-based reading and correction of answer sheets. This study implements a computer vision-based OMR system that utilizes a blank answer sheet template as a reference for identifying mark changes in answer-choice areas. The system was developed using the Python programming language and the OpenCV library through several stages, including preprocessing, grayscale conversion, adaptive thresholding, contour detection, perspective correction, image alignment with the template, answer-area segmentation, answer identification, answer correction based on the answer key, and score calculation. The system was designed to recognize six output classes: A, B, C, D, E, and BLANK. Evaluation was conducted using 50 answer-sheet images, each containing 60 questions with five answer alternatives, resulting in 3,000 answer decisions and 15,000 answer-choice areas being analyzed. All 50 answer sheets were successfully processed during the validation stage. The evaluation results showed that 2,863 predictions were correctly identified and 137 predictions were incorrect, yielding an accuracy of 95.43%, macro precision of 95.62%, macro recall of 94.53%, and macro F1-score of 94.92%. Specific evaluation of the BLANK class achieved a precision of 100%, recall of 85.71%, and F1-score of 92.31%. These results indicate that the computer vision- and template-based OMR approach can be used to automatically read and correct answer sheets with good performance. Nevertheless, 18 answers that were actually blank were identified as one of the A–E options. This condition indicates that visual variations in answer areas can still affect the identification process; therefore, optimization of the template-comparison mechanism and testing under more diverse image conditions are required to improve the robustness of the system.

Keywords: Optical Mark Recognition; Computer Vision; Computer Answer Sheet; Image Processing; Automatic Correction

1. PENDAHULUAN

Pendidikan merupakan salah satu aspek penting dalam kehidupan manusia yang berperan dalam meningkatkan pengetahuan, keterampilan, dan kemampuan peserta didik. Dalam proses pendidikan, kegiatan pembelajaran perlu disertai dengan evaluasi untuk mengetahui tingkat pemahaman peserta didik terhadap materi yang telah diberikan. Salah satu bentuk evaluasi yang banyak digunakan adalah ujian dengan format pilihan ganda. Bentuk ujian tersebut memungkinkan proses penilaian dilakukan secara terstruktur dan dapat digunakan untuk mengukur kemampuan peserta didik dalam jumlah yang relatif banyak dalam waktu yang terbatas.

Ujian merupakan bagian dari proses evaluasi pembelajaran yang digunakan untuk mengetahui tingkat pencapaian peserta didik setelah mengikuti kegiatan pembelajaran. Melalui pelaksanaan ujian, guru dapat memperoleh informasi mengenai tingkat pemahaman peserta didik terhadap materi yang telah disampaikan. Salah satu bentuk ujian yang masih banyak diterapkan adalah ujian tertulis menggunakan Lembar Jawaban manual. Pada Lembar Jawab manual, peserta didik





memberikan tanda pada salah satu pilihan jawaban yang tersedia. Penggunaan Lembar Jawab dapat membantu proses pemeriksaan karena format jawaban telah dibuat secara terstruktur dan memiliki posisi pilihan yang telah ditentukan [1].

Meskipun demikian, pemeriksaan Lembar Jawab dalam jumlah banyak masih dapat menjadi permasalahan apabila dilakukan secara manual. Pemeriksaan secara manual membutuhkan waktu dan tenaga yang cukup besar, terutama ketika jumlah peserta dan soal yang harus diperiksa semakin banyak[2]. Selain itu, proses pemeriksaan manual berpotensi menimbulkan kesalahan akibat ketidaktekelitian, kelelahan, maupun inkonsistensi dalam membaca tanda yang diberikan peserta didik [3]. Kondisi tersebut menunjukkan perlunya suatu sistem yang dapat membantu melakukan pemeriksaan Lembar Jawab secara otomatis sehingga proses koreksi dapat dilakukan dengan lebih cepat dan terstruktur [4].

Salah satu teknologi yang dapat digunakan untuk mengatasi permasalahan tersebut adalah Optical Mark Recognition (OMR). OMR merupakan teknologi yang digunakan untuk mengenali tanda atau mark yang terdapat pada dokumen tertentu secara otomatis. Teknologi ini banyak diterapkan pada formulir, kuesioner, survei, dan lembar ujian pilihan ganda karena mampu membaca tanda pada area yang telah ditentukan [5]. Penerapan OMR pada Lembar Jawab dapat membantu mengidentifikasi pilihan jawaban peserta didik dan mengubah hasil pembacaan tersebut menjadi data yang dapat diproses oleh sistem.

Perkembangan teknologi computer vision dan pengolahan citra digital memberikan peluang untuk mengembangkan sistem OMR berbasis perangkat lunak. Salah satu pustaka yang banyak digunakan dalam pengolahan citra adalah OpenCV (OpenSource Computer Vision Library) yang menyediakan berbagai fungsi untuk melakukan manipulasi dan analisis citra, seperti konversi citra, thresholding, deteksi kontur, transformasi perspektif, dan proses pengolahan citra lainnya [6]. Bahasa pemrograman Python juga dapat digunakan dalam pengembangan sistem tersebut karena memiliki berbagai pustaka yang mendukung implementasi pengolahan citra dan computer vision [7]. Dengan memanfaatkan teknologi tersebut, proses pembacaan Lembar Jawab dapat dilakukan menggunakan citra yang diperoleh dari perangkat kamera atau pemindai tanpa harus bergantung sepenuhnya pada mesin OMR khusus.

Penelitian OMR berbasis computer vision telah banyak memanfaatkan teknik pengolahan citra untuk mendeteksi struktur lembar jawaban dan mengenali tanda pada area pilihan. Yusuf, Pradana, dan Susanto [8]. mengembangkan sistem OMR berbasis OpenCV yang menerapkan adaptive thresholding, deteksi kontur, transformasi perspektif, dan segmentasi area jawaban berdasarkan analisis piksel. Hasil penelitian menunjukkan bahwa pendekatan tersebut mampu mencapai akurasi tinggi pada arsip yang jelas, tetapi kinerjanya menurun ketika citra mengandung coretan yang signifikan. Temuan tersebut menunjukkan bahwa pendekatan berbasis pengolahan citra dapat memberikan hasil yang baik pada struktur Lembar Jawab yang terkontrol, tetapi karakteristik visual seperti noise, coretan, dan variasi kualitas citra tetap dapat memengaruhi proses identifikasi [9]. Penelitian lain mengenai OMR pada skala besar juga menunjukkan bahwa metode berbasis pengolahan citra dapat digunakan untuk mengekstraksi area jawaban dan melakukan pengenalan tanda, tetapi kondisi citra dan variasi struktur lembar tetap menjadi faktor yang perlu diperhatikan dalam pengembangan sistem OMR [10].

Selain pendekatan berbasis pengolahan citra, penelitian OMR berkembang menuju penggunaan Machine Learning dan Deep Learning untuk mempelajari karakteristik visual tanda jawaban. Mondal et al. [11] mengembangkan OMRNet, yaitu model Convolutional Neural Network (CNN) berbasis transfer learning yang dirancang untuk mengklasifikasikan area jawaban ke dalam kondisi confirmed, crossed-out, dan empty. Pendekatan tersebut menunjukkan bahwa karakteristik tanda dapat dipelajari secara langsung dari data sehingga sistem tidak sepenuhnya bergantung pada aturan berbasis intensitas atau parameter pengolahan citra yang ditentukan secara manual. Selain CNN, pendekatan object detection juga mulai diterapkan pada OMR. Salah satu penelitian menggunakan YOLOv8n untuk mendeteksi dan mengklasifikasikan tanda pada Lembar Jawab sehingga proses pengenalan tidak hanya bergantung pada posisi area jawaban yang telah ditentukan sebelumnya [12]. Perkembangan tersebut menunjukkan bahwa Machine Learning dan Deep Learning memiliki potensi untuk menangani variasi visual tanda dan struktur Lembar Jawab yang lebih beragam.

Meskipun Machine Learning dan Deep Learning menawarkan kemampuan untuk mempelajari karakteristik visual tanda secara otomatis, pendekatan tersebut memiliki kebutuhan yang berbeda dibandingkan metode computer vision berbasis template. Pendekatan Machine Learning dan Deep Learning memerlukan dataset yang representatif dan, pada pendekatan supervised, data berlabel untuk proses pelatihan dan validasi model. Pengembangan model juga melibatkan pemilihan arsitektur, konfigurasi parameter, proses pelatihan, dan evaluasi model. Sebaliknya, pendekatan computer vision berbasis template tidak memerlukan proses pelatihan model karena memanfaatkan struktur geometris dan citra referensi yang telah ditentukan. Pendekatan ini lebih sesuai untuk Lembar Jawab dengan layout yang terkontrol, posisi area jawaban yang konsisten, dan jumlah serta susunan pilihan yang telah diketahui. Konsekuensinya, pendekatan template memiliki ketergantungan yang lebih tinggi terhadap kesesuaian struktur dan alignment citra, sedangkan pendekatan Machine Learning dan Deep Learning berpotensi memberikan fleksibilitas yang lebih tinggi ketika tersedia data pelatihan dengan variasi kondisi yang memadai. Oleh karena itu, penelitian ini tidak menempatkan pendekatan template sebagai pengganti Machine Learning atau Deep Learning, tetapi mengevaluasi implementasi pendekatan computer vision berbasis template pada struktur Lembar Jawab yang terkontrol.

Salah satu permasalahan penting dalam sistem OMR adalah membedakan jawaban yang benar-benar kosong dari area yang mengandung tanda atau elemen visual lain pada Lembar Jawab. Pada Lembar Jawab yang digunakan dalam penelitian ini, setiap area pilihan memiliki elemen cetak seperti garis tabel, nomor soal, dan karakter pilihan A–E. Apabila identifikasi hanya didasarkan pada intensitas atau jumlah piksel gelap pada suatu area, elemen cetak tersebut dapat menghasilkan respons yang menyerupai tanda jawaban. Selain itu, noise citra, perubahan pencahayaan, ketebalan arsip, dan ketidakteraturan tanda dapat mengubah karakteristik piksel pada area pilihan. Kondisi tersebut dapat menyebabkan



sistem memberikan prediksi A–E pada area yang sebenarnya tidak memiliki jawaban. Permasalahan ini menunjukkan bahwa keberhasilan sistem OMR tidak hanya ditentukan oleh kemampuan mengenali tanda yang terisi, tetapi juga oleh kemampuan membedakan perubahan yang berasal dari tanda peserta didik dengan informasi visual yang memang telah terdapat pada Lembar Jawab.

Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan template Lembar Jawab kosong sebagai citra referensi sebelum pemberian tanda oleh peserta didik. Template merepresentasikan kondisi dasar dokumen yang telah memiliki seluruh elemen cetak, seperti garis tabel, nomor soal, dan karakter pilihan A–E. Dengan menggunakan kondisi tersebut sebagai referensi, proses identifikasi tidak hanya didasarkan pada nilai intensitas absolut suatu area, tetapi dapat mempertimbangkan perubahan visual antara Lembar Jawab yang telah diisi dan kondisi awalnya. Dengan demikian, elemen visual yang telah terdapat pada template dapat digunakan sebagai baseline dalam proses perbandingan, sedangkan perubahan yang muncul pada area pilihan menjadi informasi yang digunakan untuk menentukan keberadaan tanda jawaban. Pendekatan tersebut diharapkan dapat membantu sistem membedakan tanda peserta didik dari elemen cetak yang terdapat pada Lembar Jawab, khususnya dalam proses identifikasi kondisi KOSONG.

Pembandingan antara Lembar Jawab dan template memerlukan kesesuaian posisi spasial agar area yang dibandingkan berada pada koordinat yang sama. Citra yang diperoleh melalui kamera dapat mengalami pergeseran posisi, perubahan orientasi, dan distorsi perspektif sehingga area pilihan pada citra aktual tidak selalu berada tepat pada posisi yang sama dengan template [13]. Apabila pembandingan dilakukan tanpa penyalarsan, perbedaan posisi tersebut dapat menghasilkan perubahan piksel yang bukan berasal dari tanda jawaban. Oleh karena itu, penelitian ini menggunakan Enhanced Correlation Coefficient (ECC) pada tahap penyalarsan citra sebelum proses pembandingan template. ECC digunakan untuk mengoptimalkan kesesuaian antara citra Lembar Jawab dan template berdasarkan korelasi intensitas, sehingga struktur visual keduanya dapat ditempatkan pada posisi yang lebih sesuai [14]. Dalam penelitian ini, ECC berfungsi sebagai tahap alignment, sedangkan template kosong berfungsi sebagai referensi kondisi dasar dokumen. Dengan pemisahan fungsi tersebut, proses identifikasi jawaban dilakukan setelah citra Lembar Jawab memiliki kesesuaian spasial yang diperlukan untuk membandingkan perubahan pada area pilihan.=

Berdasarkan penelitian terdahulu [15], pendekatan OMR berbasis computer vision memiliki keunggulan dalam kesederhanaan implementasi dan dapat diterapkan tanpa proses pelatihan model, sedangkan pendekatan Machine Learning dan Deep Learning memberikan peluang untuk mempelajari karakteristik tanda dan menangani variasi visual yang lebih kompleks. Namun, kedua pendekatan tersebut memiliki karakteristik dan kebutuhan implementasi yang berbeda. Pada Lembar Jawab dengan struktur yang telah ditentukan, pendekatan template-based masih memiliki peluang untuk digunakan karena lokasi dan susunan area jawaban dapat didefinisikan secara eksplisit. Permasalahan yang perlu dikaji adalah bagaimana pendekatan tersebut dapat membedakan perubahan akibat tanda peserta didik dari elemen cetak yang telah terdapat pada Lembar Jawab, terutama ketika harus mengenali kondisi KOSONG. Selain itu, kesesuaian spasial antara citra aktual dan template menjadi faktor penting karena pergeseran atau perubahan orientasi dapat memengaruhi hasil pembandingan. Oleh karena itu, penelitian ini berfokus pada implementasi OMR berbasis computer vision dengan template Lembar Jawab kosong dan penyalarsan citra menggunakan ECC untuk melakukan identifikasi pilihan A, B, C, D, E, dan KOSONG serta mengevaluasi kinerjanya pada Lembar Jawab dengan struktur yang terkontrol.

2. METODE PENELITIAN

2.1 Preprocessing

Penelitian ini menggunakan pendekatan eksperimental untuk mengimplementasikan Optical Mark Recognition (OMR) berbasis computer vision dengan memanfaatkan template Lembar Jawab kosong sebagai citra referensi. Pendekatan yang dikembangkan ditujukan untuk mengenali enam kondisi jawaban, yaitu pilihan A, B, C, D, E, dan KOSONG. Prinsip utama metode adalah membandingkan citra Lembar Jawab yang telah diisi dengan template kosong setelah kedua citra berada pada posisi spasial yang selaras [16]. Dengan pendekatan tersebut, elemen visual yang telah terdapat pada Lembar Jawab, seperti garis tabel, nomor soal, dan karakter pilihan, dapat diperlakukan sebagai bagian dari kondisi dasar dokumen sehingga perubahan visual akibat tanda peserta didik dapat dianalisis secara lebih terarah [17].

Tahapan penelitian terdiri atas akuisisi citra, preprocessing, penyalarsan citra menggunakan Enhanced Correlation Coefficient (ECC), identifikasi area jawaban, pembandingan citra Lembar Jawab dengan template kosong, penentuan kondisi jawaban berdasarkan nilai perubahan citra, serta evaluasi kinerja sistem. Tahap preprocessing digunakan untuk mempersiapkan citra agar struktur Lembar Jawab dapat diproses secara konsisten. Tahap alignment menggunakan ECC dilakukan untuk mengurangi perbedaan posisi antara citra aktual dan template akibat pergeseran atau perubahan orientasi. Setelah citra selaras, setiap area pilihan dibandingkan dengan area yang bersesuaian pada template untuk memperoleh nilai perubahan visual yang menjadi dasar identifikasi jawaban.

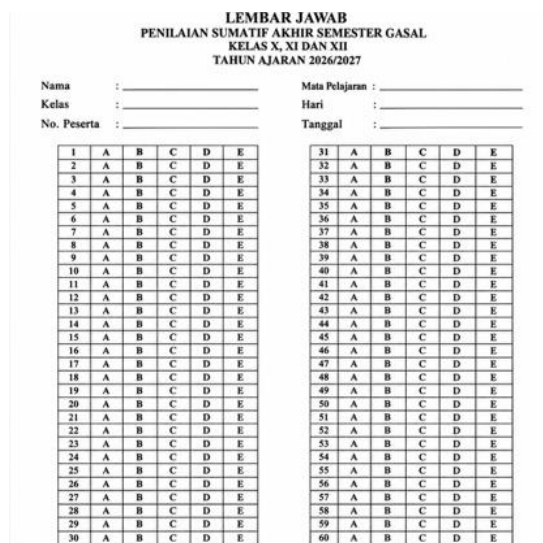
Secara konseptual, proses identifikasi dalam penelitian ini terdiri atas tiga komponen utama, yaitu template kosong sebagai representasi kondisi awal dokumen, ECC sebagai mekanisme penyalarsan spasial, dan pengukuran perubahan citra sebagai dasar pengambilan keputusan. Template tidak digunakan untuk menentukan jawaban secara langsung, tetapi berfungsi sebagai baseline terhadap kondisi visual Lembar Jawab sebelum diberikan tanda. ECC digunakan untuk memastikan bahwa area pada citra aktual dan template memiliki korespondensi spasial sebelum dilakukan pembandingan. Selanjutnya, perubahan visual pada setiap area pilihan dihitung dan digunakan untuk menentukan apakah area tersebut termasuk kondisi KOSONG atau mengandung tanda jawaban A–E.



Evaluasi dilakukan menggunakan 50 citra Lembar Jawab dengan masing-masing Lembar Jawab terdiri atas 60 soal pilihan ganda. Dengan demikian, terdapat 3.000 keputusan jawaban yang dievaluasi dan 15.000 area pilihan yang diproses. Kinerja sistem dievaluasi menggunakan accuracy, precision, recall, dan F1-score untuk mengukur kemampuan sistem dalam mengidentifikasi pilihan jawaban. Evaluasi terhadap kelas KOSONG dilakukan secara khusus untuk mengetahui kemampuan metode dalam membedakan area tanpa tanda dari area yang mengandung jawaban. Hasil evaluasi tersebut digunakan untuk menilai kemampuan pendekatan template-based dalam melakukan koreksi otomatis pada Lembar Jawab dengan struktur yang telah ditentukan.

2.2 Dataset dan Karakteristik Lembar Jawab

Dataset penelitian terdiri atas 50 citra Lembar Jawab Komputer (LJK) yang digunakan untuk menguji implementasi sistem Optical Mark Recognition (OMR). Setiap LJK memiliki 60 soal pilihan ganda dengan lima alternatif jawaban, yaitu A, B, C, D, dan E. Struktur tersebut menghasilkan 300 area pilihan jawaban pada setiap LJK, sehingga keseluruhan dataset mencakup 3.000 keputusan jawaban atau 15.000 area pilihan yang diproses oleh sistem. Gambar 1 merupakan bentuk lembar jawab yang digunakan pada penelitian ini.



LEMBAR JAWAB
PENILAIAN SUMATIF AKHIR SEMESTER GASAL
KELAS X, XI DAN XII
TAHUN AJARAN 2026/2027

Nama : _____ Mata Pelajaran : _____
Kelas : _____ Hari : _____
No. Peserta : _____ Tanggal : _____

1	A	B	C	D	E
2	A	B	C	D	E
3	A	B	C	D	E
4	A	B	C	D	E
5	A	B	C	D	E
6	A	B	C	D	E
7	A	B	C	D	E
8	A	B	C	D	E
9	A	B	C	D	E
10	A	B	C	D	E
11	A	B	C	D	E
12	A	B	C	D	E
13	A	B	C	D	E
14	A	B	C	D	E
15	A	B	C	D	E
16	A	B	C	D	E
17	A	B	C	D	E
18	A	B	C	D	E
19	A	B	C	D	E
20	A	B	C	D	E
21	A	B	C	D	E
22	A	B	C	D	E
23	A	B	C	D	E
24	A	B	C	D	E
25	A	B	C	D	E
26	A	B	C	D	E
27	A	B	C	D	E
28	A	B	C	D	E
29	A	B	C	D	E
30	A	B	C	D	E
31	A	B	C	D	E
32	A	B	C	D	E
33	A	B	C	D	E
34	A	B	C	D	E
35	A	B	C	D	E
36	A	B	C	D	E
37	A	B	C	D	E
38	A	B	C	D	E
39	A	B	C	D	E
40	A	B	C	D	E
41	A	B	C	D	E
42	A	B	C	D	E
43	A	B	C	D	E
44	A	B	C	D	E
45	A	B	C	D	E
46	A	B	C	D	E
47	A	B	C	D	E
48	A	B	C	D	E
49	A	B	C	D	E
50	A	B	C	D	E
51	A	B	C	D	E
52	A	B	C	D	E
53	A	B	C	D	E
54	A	B	C	D	E
55	A	B	C	D	E
56	A	B	C	D	E
57	A	B	C	D	E
58	A	B	C	D	E
59	A	B	C	D	E
60	A	B	C	D	E

Gambar 1. Lembar jawab

Setiap citra LJK memiliki struktur yang relatif tetap, dengan area soal 1–30 berada pada bagian kiri dan soal 31–60 pada bagian kanan. Setiap soal memiliki lima area pilihan yang merepresentasikan alternatif A–E. Struktur yang terkontrol tersebut digunakan sebagai dasar penentuan posisi area jawaban pada tahap segmentasi dan perbandingan template. Dengan posisi area yang telah diketahui, sistem dapat melakukan perbandingan antara setiap area pada LJK aktual dengan area yang bersesuaian pada template LJK kosong.

Dalam proses pengenalan, sistem membedakan enam kelas kondisi jawaban, yaitu A, B, C, D, E, dan KOSONG. Kelas A–E menunjukkan bahwa salah satu alternatif jawaban diberikan tanda oleh peserta didik, sedangkan kelas KOSONG menunjukkan bahwa tidak terdapat tanda jawaban pada area soal. Kelas KOSONG diperlakukan sebagai kelas tersendiri karena kemampuan membedakan area yang benar-benar tidak diisi dari elemen cetak pada LJK merupakan salah satu permasalahan utama yang dikaji dalam penelitian ini.

Sebelum digunakan dalam evaluasi, seluruh citra LJK diproses melalui tahapan pemeriksaan dan validasi terhadap struktur lembar. Dari 50 citra yang diuji, seluruh citra berhasil diproses oleh sistem sehingga diperoleh 50 citra valid dan tidak terdapat citra yang gagal diproses. Dengan demikian, seluruh 3.000 keputusan jawaban dan 15.000 area pilihan dapat digunakan dalam evaluasi kinerja sistem.

Dataset digunakan untuk mengevaluasi kemampuan metode dalam mengenali pilihan jawaban sekaligus mengidentifikasi kondisi KOSONG. Evaluasi dilakukan pada tingkat keputusan jawaban dengan membandingkan hasil prediksi sistem terhadap label jawaban yang telah ditentukan. Selanjutnya, kinerja sistem dihitung menggunakan accuracy, precision, recall, dan F1-score. Evaluasi khusus terhadap kelas KOSONG dilakukan untuk mengetahui kemampuan metode dalam mendeteksi tidak adanya tanda pada area jawaban, karena kesalahan pada kelas tersebut dapat menyebabkan area kosong dianggap sebagai salah satu pilihan A–E.

2.3 Preprocessing dan Grayscale

Preprocessing merupakan tahap awal yang dilakukan untuk mempersiapkan citra Lembar Jawab sebelum memasuki proses analisis struktur dan identifikasi jawaban [18]. Citra diperoleh menggunakan kamera sehingga karakteristik citra dapat dipengaruhi oleh kondisi pencahayaan, posisi pengambilan, orientasi dokumen, dan kualitas citra. Oleh karena itu, preprocessing dilakukan untuk menghasilkan representasi citra yang lebih sesuai dengan tahapan pengolahan berikutnya.

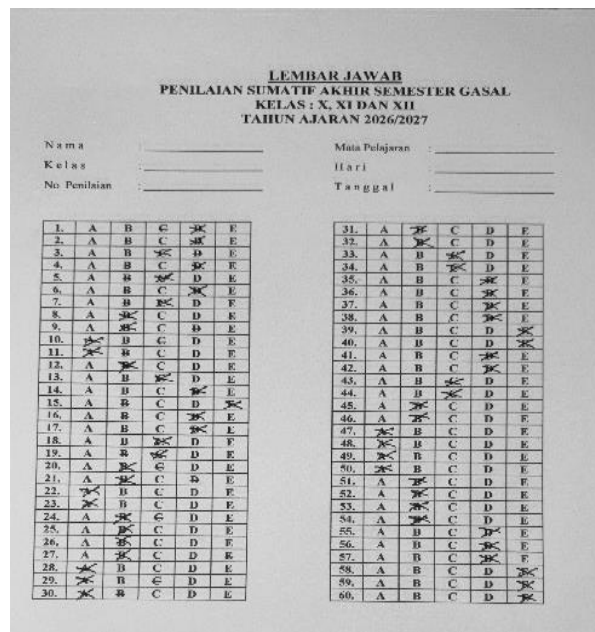
Pada tahap awal, citra Lembar Jawab yang diperoleh dalam format RGB dikonversi menjadi citra grayscale. Citra RGB memiliki tiga kanal warna, sedangkan proses identifikasi struktur Lembar Jawab pada penelitian ini terutama memanfaatkan informasi intensitas. Konversi tersebut menghasilkan satu nilai intensitas untuk setiap piksel sehingga representasi citra menjadi lebih sederhana dan dapat digunakan sebagai masukan pada proses adaptive thresholding [19]. Secara matematis, citra grayscale dapat direpresentasikan sebagai fungsi intensitas:

$$I_g(x, y) = 0.299R(x, y) + 0.587G(x, y) + 0.114B(x, y) \quad (1)$$

Dengan $(R(x, y))$, $(G(x, y))$, dan $(B(x, y))$ masing-masing merupakan nilai intensitas kanal merah, hijau, dan biru pada koordinat $((x, y))$, sedangkan $(I_g(x, y))$ merupakan nilai intensitas grayscale. Hasil konversi menghasilkan citra satu kanal dengan rentang intensitas yang digunakan sebagai masukan untuk proses pengolahan citra berikutnya.

Pada penelitian ini, konversi grayscale tidak digunakan sebagai dasar akhir untuk menentukan jawaban A, B, C, D, E, atau KOSONG. Tahap ini hanya menghasilkan representasi intensitas yang diperlukan untuk proses ekstraksi struktur Lembar Jawab. Penentuan kondisi jawaban dilakukan pada tahap berikutnya melalui proses perbandingan antara citra Lembar Jawab dan template kosong setelah citra diselaraskan.

Pemisahan fungsi tersebut diperlukan karena intensitas gelap pada citra tidak selalu menunjukkan adanya tanda dari peserta didik. Elemen cetak pada Lembar Jawab, seperti garis tabel, nomor soal, dan karakter pilihan A–E, juga dapat menghasilkan nilai intensitas yang berbeda dari latar belakang. Oleh karena itu, informasi grayscale digunakan sebagai representasi awal citra, sedangkan diferensiasi antara elemen cetak dan tanda jawaban dilakukan pada tahap perbandingan terhadap template kosong. Gambar 2 merupakan contoh lembar jawab yang sudah diubah menjadi citra grayscale.



Gambar 2. Citra grayscale

Setelah citra dikonversi menjadi grayscale, citra diproses menggunakan adaptive thresholding untuk membantu memperoleh struktur geometris Lembar Jawab. Hasil thresholding selanjutnya digunakan sebagai informasi pendukung dalam proses deteksi kontur dan koreksi perspektif. Dengan demikian, preprocessing dan grayscale berfungsi sebagai tahap persiapan citra, sedangkan proses utama identifikasi tanda jawaban dilakukan melalui penyelarasan citra dan perbandingan dengan template yang dijelaskan pada tahapan metode berikutnya.

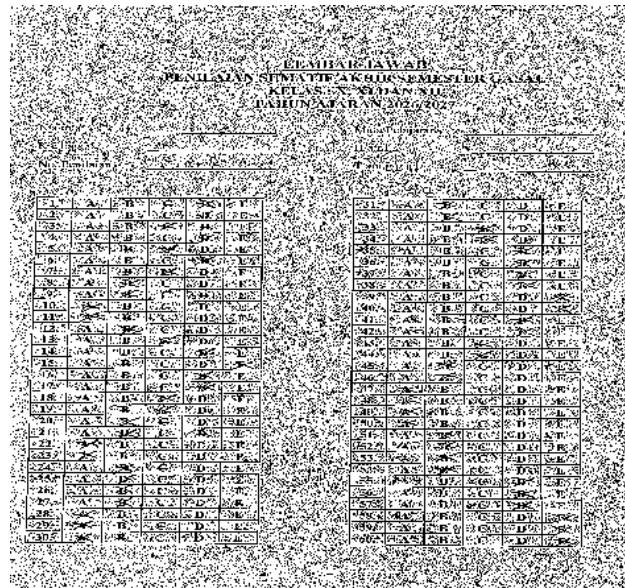
2.4 Adaptive Thresholding

Setelah citra Lembar Jawab dikonversi menjadi grayscale, tahap berikutnya adalah adaptive thresholding untuk menghasilkan representasi biner yang digunakan dalam proses ekstraksi struktur dokumen [20]. Tahap ini diperlukan karena citra yang diperoleh melalui kamera dapat memiliki distribusi pencahayaan yang tidak seragam pada seluruh bagian Lembar Jawab. Penggunaan nilai ambang yang sama untuk seluruh citra berpotensi menghasilkan bagian dokumen yang terlalu gelap atau terlalu terang, sehingga struktur Lembar Jawab sulit diperoleh secara konsisten [21].

Secara umum, proses thresholding dapat dinyatakan sebagai:

$$B(x, y) = \begin{cases} 1, & I(x, y) < T(x, y) \\ 0, & I(x, y) \geq T(x, y) \end{cases} \quad (2)$$

Dengan $(I(x, y))$ merupakan intensitas piksel pada koordinat (x, y) , $(T(x, y))$ merupakan nilai ambang lokal, dan $(B(x, y))$ merupakan nilai piksel pada citra biner. Gambar 3 merupakan citra yang dihasilkan dari proses adaptive thresholding.



Gambar 3. Citra adaptive thresholding

Pada penelitian ini, citra biner yang dihasilkan dari adaptive thresholding digunakan untuk membantu memperoleh struktur geometris Lembar Jawab, terutama dalam proses pendeteksian batas tabel dan area dokumen. Dengan demikian, nilai ambang ($T(x,y)$) pada tahap ini digunakan untuk ekstraksi struktur, bukan sebagai nilai ambang untuk menentukan apakah suatu area pilihan merupakan jawaban terisi atau KOSONG.

Pemisahan fungsi tersebut penting karena elemen cetak pada Lembar Jawab, seperti garis tabel, nomor soal, dan karakter pilihan A–E, juga dapat menghasilkan piksel dengan karakteristik intensitas yang berbeda dari latar belakang. Apabila hasil thresholding digunakan secara langsung sebagai dasar identifikasi jawaban, elemen cetak tersebut berpotensi dianggap sebagai tanda peserta didik.

Oleh karena itu, adaptive thresholding dalam penelitian ini ditempatkan sebagai tahap pendukung sebelum proses deteksi kontur dan koreksi perspektif. Setelah struktur Lembar Jawab diperoleh dan citra berada pada posisi yang sesuai, identifikasi jawaban dilakukan melalui mekanisme yang berbeda, yaitu penyesuaian citra terhadap template kosong dan pembandingan perubahan visual pada area pilihan. Formulasi mengenai perbedaan antara citra Lembar Jawab dan template serta nilai ambang untuk menentukan kondisi A–E atau KOSONG dijelaskan pada tahapan identifikasi jawaban.

2.5 Deteksi Kontur dan Koreksi Perspektif

Citra Lembar Jawab yang diperoleh menggunakan kamera dapat mengalami kemiringan dan perubahan perspektif akibat posisi kamera yang tidak selalu tegak lurus terhadap bidang dokumen. Kondisi tersebut dapat menyebabkan perubahan posisi dan bentuk area jawaban sehingga koordinat setiap pilihan tidak berada pada posisi yang konsisten. Oleh karena itu, diperlukan proses deteksi struktur Lembar Jawab dan koreksi perspektif sebelum dilakukan segmentasi area jawaban.

Deteksi kontur digunakan untuk memperoleh batas geometris objek pada citra hasil thresholding. Pada penelitian ini, kontur dimanfaatkan untuk menemukan struktur tabel pada Lembar Jawab yang menjadi acuan dalam menentukan area pemrosesan. Informasi batas geometris tersebut digunakan untuk memisahkan bagian Lembar Jawab yang akan dianalisis dari bagian citra lainnya.

Setelah struktur tabel diperoleh, dilakukan koreksi perspektif untuk mengubah citra Lembar Jawab yang mengalami kemiringan menjadi citra dengan posisi yang lebih teratur. Proses ini menggunakan transformasi perspektif berdasarkan empat titik sudut area yang telah ditentukan. Secara matematis, transformasi perspektif dapat dinyatakan sebagai:

$$\begin{aligned}x' &= h_{11}x + h_{12}y + h_{13} \\y' &= h_{21}x + h_{22}y + h_{23} \\w' &= h_{31}x + h_{32}y + h_{33}\end{aligned}\tag{3}$$

Koordinat hasil transformasi kemudian dinormalisasi menggunakan persamaan:

$$\begin{aligned}x^* &= x' / w' \\y^* &= y' / w'\end{aligned}\tag{4}$$

Dengan x dan y merupakan koordinat piksel pada citra awal, x^* dan y^* merupakan koordinat setelah koreksi perspektif, sedangkan h_{11} sampai h_{33} merupakan parameter transformasi perspektif. Nilai w' digunakan untuk melakukan normalisasi koordinat hasil transformasi.



Koreksi perspektif bertujuan untuk menghasilkan posisi Lembar Jawab yang lebih konsisten sehingga struktur area jawaban dapat dipetakan berdasarkan tata letak template yang digunakan. Dengan demikian, proses ini berfungsi sebagai tahap normalisasi geometris dan belum digunakan untuk menentukan apakah suatu pilihan jawaban terisi atau kosong.

Setelah koreksi perspektif selesai, Lembar Jawab dipisahkan menjadi dua bagian utama. Tabel kiri digunakan untuk memproses soal nomor 1–30, sedangkan tabel kanan digunakan untuk memproses soal nomor 31–60. Hasil koreksi perspektif selanjutnya digunakan sebagai masukan pada proses segmentasi area jawaban.

Setelah proses koreksi perspektif, Lembar Jawab dipisahkan menjadi dua bagian utama, yaitu tabel kiri dan tabel kanan. Tabel kiri digunakan untuk memproses soal nomor 1–30, sedangkan tabel kanan digunakan untuk memproses soal nomor 31–60.

2.6 Segmentasi Area Jawaban

Setelah proses koreksi perspektif selesai, citra Lembar Jawab berada pada posisi yang lebih teratur sehingga area pilihan jawaban dapat dipetakan berdasarkan struktur template. Tahap segmentasi dilakukan untuk membagi Lembar Jawab menjadi sejumlah region of interest (ROI) yang masing-masing merepresentasikan satu pilihan jawaban.

Struktur Lembar Jawab terdiri atas 60 soal dengan lima alternatif jawaban, yaitu A, B, C, D, dan E. Lembar Jawab dibagi menjadi dua tabel, yaitu tabel kiri yang memuat soal nomor 1–30 dan tabel kanan yang memuat soal nomor 31–60. Setiap tabel terdiri atas 30 soal dengan lima area pilihan sehingga masing-masing tabel menghasilkan 150 area pilihan.

Dengan demikian, jumlah area pilihan yang diperoleh dari satu Lembar Jawab adalah:

$$30 \text{ soal} \times 5 \text{ pilihan} \times 2 \text{ tabel} = 300 \text{ area pilihan}$$

Setiap area pilihan direpresentasikan sebagai ROI yang dikaitkan dengan nomor soal dan alternatif jawaban tertentu. ROI untuk soal ke-q dan pilihan ke-k dapat dinyatakan sebagai $R(q,k)$, dengan q menunjukkan nomor soal dan k menunjukkan alternatif jawaban A, B, C, D, atau E.

Proses segmentasi menghasilkan kelompok area yang terdiri atas lima ROI untuk setiap nomor soal. Sebagai contoh, soal nomor 1 memiliki lima ROI yang masing-masing merepresentasikan pilihan A, B, C, D, dan E. Pola yang sama diterapkan hingga soal nomor 60.

Segmentasi pada tahap ini tidak digunakan untuk menentukan apakah suatu pilihan terisi atau kosong. Tahap segmentasi hanya bertujuan menetapkan batas dan posisi masing-masing area pilihan agar proses perbandingan citra dapat dilakukan secara konsisten. Setiap ROI hasil segmentasi selanjutnya digunakan sebagai masukan pada tahap penyalarsan citra dan perbandingan terhadap template Lembar Jawab kosong.

Dengan menggunakan 50 Lembar Jawab, proses segmentasi menghasilkan 15.000 area pilihan yang menjadi objek analisis, yaitu $50 \text{ Lembar Jawab} \times 60 \text{ soal} \times 5 \text{ pilihan}$. Setiap area tersebut kemudian dianalisis untuk menentukan karakteristik perubahan citra akibat adanya tanda jawaban.

2.7 Penyalarsan Citra Menggunakan ECC

Setelah proses segmentasi area jawaban, citra Lembar Jawab perlu diselaraskan dengan template Lembar Jawab kosong. Penyalarsan dilakukan karena masih dapat terdapat perbedaan posisi antara citra hasil pemotretan dengan template, meskipun sebelumnya telah dilakukan koreksi perspektif [22]. Perbedaan tersebut dapat berupa pergeseran posisi, perubahan orientasi, maupun ketidaktepatan kecil pada posisi area jawaban.

Pada penelitian ini, penyalarsan citra dilakukan menggunakan Enhanced Correlation Coefficient (ECC). Metode ECC digunakan untuk memperoleh transformasi geometris yang menghasilkan kesesuaian posisi antara citra Lembar Jawab yang diuji dengan citra template kosong. Dengan penyalarsan tersebut, posisi area jawaban pada kedua citra dapat berada pada koordinat yang relatif sama sehingga proses perbandingan citra dapat dilakukan secara lebih konsisten.

Secara umum, proses penyalarsan ECC bertujuan memaksimalkan tingkat korelasi antara citra masukan dan template. Nilai korelasi dapat dinyatakan sebagai:

$$ECC = \text{correlation}(I, T) \quad (5)$$

Dengan I merupakan citra Lembar Jawab yang diuji setelah melalui tahap preprocessing dan koreksi perspektif, sedangkan T merupakan citra template Lembar Jawab kosong. Proses optimasi dilakukan dengan mencari transformasi geometris yang menghasilkan nilai korelasi maksimum antara kedua citra.

Transformasi tersebut digunakan untuk menghasilkan citra Lembar Jawab yang telah terselaraskan terhadap template. Secara konseptual, hasil penyalarsan dapat dinyatakan sebagai:

$$I_{\text{aligned}} = \text{Warp}(I, W) \quad (6)$$

Dengan I_{aligned} merupakan citra Lembar Jawab setelah penyalarsan, I merupakan citra masukan, dan W merupakan parameter transformasi hasil optimasi ECC. Hasil penyalarsan digunakan sebagai dasar untuk membandingkan setiap ROI pada Lembar Jawab dengan ROI yang bersesuaian pada template kosong. Dengan demikian, tahap ECC berfungsi untuk mengurangi pengaruh perbedaan posisi antara citra pengujian dan template sebelum dilakukan analisis terhadap perubahan visual pada area pilihan jawaban.

Penyalarsan menggunakan ECC tidak digunakan untuk menentukan pilihan A, B, C, D, E, atau KOSONG secara langsung. Tahap ini hanya memastikan bahwa area pada Lembar Jawab dan area yang bersesuaian pada template berada



pada posisi yang relatif sama. Penentuan status jawaban dilakukan pada tahap berikutnya melalui pengukuran perbedaan antara ROI Lembar Jawab dan ROI template.

2.8 Perbandingan Template dan Formulasi Diferensiasi Tinta

Setelah citra Lembar Jawab diselaraskan terhadap template menggunakan ECC, tahap berikutnya adalah melakukan perbandingan antara area pilihan pada Lembar Jawab dengan area yang bersesuaian pada template Lembar Jawab kosong. Template kosong digunakan sebagai baseline yang merepresentasikan kondisi Lembar Jawab sebelum diberikan tanda oleh peserta didik. Dengan pendekatan tersebut, elemen cetak seperti garis tabel, nomor soal, dan karakter pilihan A–E yang telah terdapat pada template tidak secara langsung dianggap sebagai tanda jawaban.

Perbandingan dilakukan pada setiap region of interest (ROI) hasil segmentasi. Untuk setiap nomor soal q dan pilihan k , terdapat ROI Lembar Jawab yang telah diselaraskan dan ROI template yang bersesuaian. Perbedaan antara kedua ROI dihitung berdasarkan perubahan nilai piksel pada posisi yang sama.

Perbedaan intensitas pada setiap piksel dapat dinyatakan sebagai:

$$D(x,y) = |I_{\text{aligned}}(x,y) - T(x,y)| \quad (7)$$

Dengan $I_{\text{aligned}}(x,y)$ merupakan nilai intensitas piksel pada citra Lembar Jawab yang telah diselaraskan, sedangkan $T(x,y)$ merupakan nilai intensitas piksel pada template kosong pada koordinat yang sama. Nilai $D(x,y)$ menunjukkan besarnya perubahan intensitas antara citra Lembar Jawab dan template.

Untuk membedakan perubahan yang dianggap signifikan dari perubahan kecil akibat noise atau perbedaan pencahayaan, digunakan nilai ambang perubahan piksel δ . Piksel dikategorikan sebagai piksel perubahan apabila:

$$M(x,y) = 1, \text{ jika } D(x,y) \geq \delta \quad (8)$$

$$M(x,y) = 0, \text{ jika } D(x,y) < \delta$$

Dengan $M(x,y)$ merupakan mask perubahan dan δ merupakan ambang perubahan intensitas piksel. Dengan demikian, hanya piksel yang memiliki perbedaan intensitas minimal sebesar δ yang dihitung sebagai perubahan. Selanjutnya, jumlah piksel perubahan pada setiap ROI dibandingkan dengan jumlah seluruh piksel pada ROI. Nilai respons atau rasio perubahan suatu area dapat dihitung menggunakan:

$$S(q,k) = \Sigma M(x,y) / N \quad (9)$$

Dengan $S(q,k)$ merupakan skor perubahan pada soal q dan pilihan k , $\Sigma M(x,y)$ merupakan jumlah piksel yang memenuhi kriteria perubahan, dan N merupakan jumlah seluruh piksel pada ROI. Nilai $S(q,k)$ berada pada rentang 0 sampai 1. Nilai yang semakin mendekati 0 menunjukkan bahwa kondisi ROI pada Lembar Jawab relatif sama dengan template kosong. Sebaliknya, nilai yang lebih besar menunjukkan adanya perubahan visual yang lebih banyak dibandingkan kondisi template.

Formulasi tersebut digunakan untuk membedakan area yang tetap kosong dari area yang memiliki tambahan tanda. Pada area KOSONG, perubahan terhadap template diharapkan relatif kecil karena tidak terdapat tanda tambahan dari peserta didik. Sebaliknya, area yang diberi tanda akan menghasilkan perubahan piksel yang lebih besar sehingga nilai $S(q,k)$ cenderung meningkat.

Perbandingan template dengan demikian tidak menggunakan intensitas absolut suatu area sebagai satu-satunya dasar keputusan. Sistem terlebih dahulu menghilangkan pengaruh elemen cetak melalui penggunaan template sebagai baseline, kemudian mengukur perubahan visual yang tersisa pada masing-masing ROI. Pendekatan ini digunakan untuk membantu membedakan tanda jawaban peserta didik dari elemen visual yang memang telah terdapat pada Lembar Jawab.

Nilai $S(q,k)$ selanjutnya digunakan pada tahap penentuan jawaban. Pilihan dengan nilai respons yang memenuhi kriteria deteksi dibandingkan dengan mathematical threshold yang ditentukan pada tahap identifikasi. Apabila tidak terdapat pilihan yang memenuhi ambang tersebut, sistem menetapkan hasil identifikasi sebagai KOSONG.

2.9 Penentuan Jawaban Berdasarkan Mathematical Threshold

Tahap penentuan jawaban dilakukan setelah diperoleh nilai respons perubahan $S(q,k)$ dari masing-masing area pilihan. Setiap nomor soal memiliki lima nilai respons yang masing-masing merepresentasikan pilihan A, B, C, D, dan E. Nilai tersebut digunakan untuk menentukan apakah terdapat tanda jawaban pada salah satu area pilihan. Untuk setiap nomor soal q , nilai respons terbesar ditentukan menggunakan:

$$S_{\max}(q) = \max\{S(q,A), S(q,B), S(q,C), S(q,D), S(q,E)\} \quad (10)$$

Nilai $S_{\max}(q)$ kemudian dibandingkan dengan mathematical threshold τ yang digunakan sebagai batas keputusan antara kondisi terisi dan kosong. Aturan keputusan ditentukan sebagai berikut:

$$\text{Jika } S_{\max}(q) \geq \tau, \text{ maka soal dinyatakan TERISI.} \quad (11)$$

$$\text{Jika } S_{\max}(q) < \tau, \text{ maka soal dinyatakan KOSONG.}$$

Apabila kondisi TERISI terpenuhi, pilihan jawaban ditentukan berdasarkan area yang memiliki nilai respons terbesar. Dengan demikian, pilihan jawaban dapat ditentukan menggunakan:



$$k^* = \operatorname{argmax}_k S(q, k) \quad (12)$$

Dengan k^* merupakan pilihan jawaban yang memiliki nilai respons terbesar di antara A, B, C, D, dan E. Berdasarkan aturan tersebut, hasil identifikasi untuk setiap soal dapat berupa salah satu dari enam kelas, yaitu A, B, C, D, E, atau KOSONG. Jika nilai respons tertinggi belum mencapai threshold τ , sistem tidak menetapkan salah satu pilihan A–E dan memberikan label KOSONG.

Penggunaan threshold τ diperlukan untuk mencegah perubahan kecil pada citra akibat noise, perbedaan pencahayaan, atau ketidaktepatan penyelarasan dianggap sebagai tanda jawaban. Dengan demikian, keputusan tidak hanya didasarkan pada pilihan yang memiliki nilai respons paling tinggi, tetapi juga mempertimbangkan apakah nilai tersebut telah melewati batas minimum yang ditetapkan. Selain nilai respons maksimum, perbedaan antara nilai tertinggi dan nilai berikutnya dapat digunakan untuk melihat tingkat pemisahan antar pilihan. Margin keputusan dapat dinyatakan sebagai:

$$\text{Margin}(q) = S1(q) - S2(q) \quad (13)$$

Dengan $S1(q)$ merupakan nilai respons tertinggi dan $S2(q)$ merupakan nilai respons tertinggi kedua pada soal q . Nilai margin yang lebih besar menunjukkan bahwa pilihan dengan nilai respons tertinggi memiliki perbedaan yang lebih jelas dibandingkan pilihan lainnya. Hasil akhir identifikasi kemudian digunakan sebagai masukan pada proses koreksi jawaban. Setiap hasil identifikasi dibandingkan dengan kunci jawaban untuk menentukan status BENAR, SALAH, atau KOSONG.

2.10 Koreksi dan Perhitungan Nilai

Tahap koreksi dilakukan setelah sistem menentukan status setiap pilihan jawaban berdasarkan nilai respons area dan mathematical threshold. Hasil identifikasi setiap soal diklasifikasikan ke dalam enam kelas, yaitu A, B, C, D, E, dan KOSONG. Kelas KOSONG diberikan apabila nilai respons tertinggi pada suatu soal tidak memenuhi nilai ambang yang telah ditentukan, sehingga perubahan piksel pada area jawaban dianggap belum cukup untuk menunjukkan adanya tanda pilihan.

Koreksi jawaban dilakukan dengan membandingkan hasil identifikasi sistem dengan kunci jawaban yang telah ditentukan sebelumnya. Untuk setiap soal, sistem menghasilkan pasangan informasi berupa nomor soal dan hasil deteksi pilihan. Jika pilihan hasil deteksi sama dengan kunci jawaban, maka jawaban dikategorikan benar. Sebaliknya, apabila pilihan yang terdeteksi berbeda dengan kunci jawaban, jawaban dikategorikan salah. Jawaban dengan status KOSONG juga dikategorikan sebagai jawaban tidak benar dalam perhitungan nilai. Jumlah jawaban benar dihitung menggunakan:

$$\text{Benar} = \sum I(y_{\text{pred}}(q) = y_{\text{true}}(q)) \quad (14)$$

Dengan $I(.)$ merupakan fungsi indikator yang bernilai 1 apabila kondisi terpenuhi dan 0 apabila kondisi tidak terpenuhi, $y_{\text{pred}}(q)$ merupakan hasil deteksi sistem pada soal ke- q , dan $y_{\text{true}}(q)$ merupakan kunci jawaban soal ke- q .

Nilai akhir peserta dihitung berdasarkan perbandingan jumlah jawaban benar terhadap jumlah seluruh soal. Persamaan yang digunakan adalah:

$$\text{Nilai} = \frac{\text{Jumlah Jawaban Benar}}{\text{Jumlah Soal}} \times 100 \quad (15)$$

Dengan jumlah soal sebanyak 60, setiap jawaban benar memberikan kontribusi sebesar $1/60$ dari keseluruhan skor. Sebagai contoh, apabila peserta memperoleh 48 jawaban benar dari 60 soal, maka nilai yang diperoleh adalah:

$$\text{Nilai} = \frac{48}{60} \times 100$$

$$\text{Nilai} = 80$$

Selain menghasilkan nilai akhir, sistem menyimpan hasil koreksi pada tingkat setiap soal. Informasi tersebut mencakup nomor soal, pilihan yang terdeteksi, status jawaban, skor respons area, margin antara kandidat tertinggi dan kandidat kedua, serta kesesuaian dengan kunci jawaban. Penyimpanan informasi pada tingkat soal memungkinkan hasil koreksi ditelusuri kembali apabila ditemukan ketidaksesuaian antara hasil deteksi dan kondisi aktual pada lembar jawaban.

Dengan demikian, tahap koreksi tidak hanya menghasilkan nilai akhir, tetapi juga menyediakan informasi rinci mengenai hasil identifikasi setiap soal. Mekanisme ini menjadi dasar untuk mengevaluasi kemampuan sistem dalam membedakan jawaban terisi, jawaban kosong, dan pilihan jawaban yang sesuai dengan kunci jawaban.

2.11 Evaluasi Kinerja

Evaluasi kinerja dilakukan untuk mengukur kemampuan sistem dalam mengenali pilihan jawaban pada lembar jawaban secara otomatis. Evaluasi dilakukan pada tingkat keputusan jawaban, yaitu dengan membandingkan hasil deteksi sistem dengan label atau kunci jawaban yang telah ditetapkan [23]. Dataset pengujian terdiri atas 50 lembar jawaban dengan 60 soal pada setiap lembar, sehingga terdapat 3.000 keputusan jawaban yang dievaluasi.

Kinerja sistem dievaluasi menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score. Penggunaan beberapa metrik tersebut diperlukan karena evaluasi tidak hanya mempertimbangkan jumlah prediksi yang



benar, tetapi juga kemampuan sistem dalam mengenali masing-masing kelas jawaban, termasuk kelas KOSONG [24]. Accuracy digunakan untuk mengukur proporsi seluruh keputusan yang berhasil diklasifikasikan dengan benar terhadap keseluruhan keputusan yang dievaluasi. Persamaan accuracy dinyatakan sebagai:

$$Accuracy = \frac{N_{Correct}}{N_{Total}} \times 100\% \quad (16)$$

Precision digunakan untuk mengukur ketepatan hasil prediksi sistem pada suatu kelas. Precision menunjukkan proporsi prediksi suatu kelas yang benar dibandingkan dengan seluruh prediksi yang diberikan sistem pada kelas tersebut. Persamaannya adalah:

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

Recall digunakan untuk mengukur kemampuan sistem dalam menemukan seluruh data yang sebenarnya termasuk ke dalam suatu kelas. Persamaannya adalah:

$$Recall = \frac{TP}{TP+FN} \quad (18)$$

F1-score merupakan nilai harmonik antara precision dan recall sehingga dapat memberikan gambaran keseimbangan antara ketepatan prediksi dan kemampuan sistem dalam menemukan kelas yang benar. Persamaannya adalah:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (19)$$

Pada persamaan tersebut, TP (True Positive) menunjukkan jumlah data yang diprediksi benar sebagai suatu kelas, FP (False Positive) menunjukkan data dari kelas lain yang salah diprediksi sebagai kelas tersebut, sedangkan FN (False Negative) menunjukkan data yang seharusnya termasuk dalam suatu kelas tetapi tidak berhasil dikenali sebagai kelas tersebut.

Evaluasi precision, recall, dan F1-score dilakukan pada setiap kelas, yaitu A, B, C, D, E, dan KOSONG. Selain evaluasi per kelas, digunakan macro average untuk memperoleh nilai rata-rata kinerja dengan memberikan bobot yang sama pada setiap kelas. Pendekatan ini penting karena kelas KOSONG memiliki karakteristik yang berbeda dari kelas pilihan A–E dan secara khusus menjadi salah satu aspek yang dievaluasi dalam penelitian. Perhitungan macro average dilakukan dengan merata-ratakan nilai metrik dari seluruh kelas. Sebagai contoh, macro precision dihitung dengan:

$$Macro\ Precision = \frac{Precision_A + Precision_B + Precision_C + Precision_D + Precision_E + Precision_{KOSONG}}{6} \quad (20)$$

Prinsip yang sama diterapkan untuk memperoleh macro recall dan macro F1-score. Selain metrik klasifikasi, evaluasi juga mempertimbangkan keberhasilan proses pemrosesan lembar jawaban. Setiap LJK dinyatakan valid apabila seluruh tahapan pemrosesan dapat dilakukan dan menghasilkan keputusan jawaban untuk seluruh 60 soal. Jumlah LJK yang berhasil diproses dan jumlah LJK yang mengalami kegagalan dicatat sebagai bagian dari evaluasi sistem.

Hasil evaluasi selanjutnya digunakan untuk menganalisis kemampuan pendekatan template-based dalam mengenali pilihan jawaban, membedakan jawaban terisi dan kosong, serta menghasilkan keputusan koreksi secara otomatis. Evaluasi pada tingkat keputusan jawaban juga memungkinkan kesalahan sistem dianalisis secara lebih terperinci berdasarkan kelas A, B, C, D, E, dan KOSONG.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Sistem OMR

Penelitian ini menghasilkan sistem Optical Mark Recognition (OMR) berbasis Computer Vision untuk melakukan identifikasi dan koreksi otomatis pada Lembar Jawab pilihan ganda. Sistem diimplementasikan menggunakan bahasa pemrograman Python dengan pustaka OpenCV sebagai komponen utama dalam proses pengolahan citra. Pendekatan yang digunakan bersifat template-based, yaitu memanfaatkan Lembar Jawab kosong sebagai citra acuan untuk membedakan elemen cetak pada Lembar Jawab dengan perubahan visual yang disebabkan oleh tanda jawaban.

Sistem menerima citra Lembar Jawab sebagai masukan. Citra tersebut diproses melalui beberapa tahapan, yaitu preprocessing, konversi grayscale, adaptive thresholding, deteksi kontur, koreksi perspektif, segmentasi area jawaban, penyalarsan citra menggunakan Enhanced Correlation Coefficient (ECC), perbandingan dengan template kosong, penentuan respons berdasarkan mathematical threshold, identifikasi jawaban, koreksi berdasarkan kunci jawaban, dan perhitungan nilai.

Struktur Lembar Jawab yang digunakan terdiri atas 60 soal dengan lima alternatif jawaban, yaitu A, B, C, D, dan E. Lembar Jawab dibagi menjadi dua bagian, yaitu tabel sebelah kiri untuk soal nomor 1–30 dan tabel sebelah kanan untuk soal nomor 31–60. Setiap soal memiliki lima area pilihan sehingga terdapat 300 area pilihan yang dianalisis pada setiap Lembar Jawab.

Pada tahap identifikasi, sistem menghasilkan enam kemungkinan kelas, yaitu A, B, C, D, E, dan KOSONG. Kelas KOSONG digunakan untuk menunjukkan kondisi ketika tidak terdapat perubahan visual yang cukup kuat pada area



Tampilan sistem dirancang dalam satu halaman sehingga proses mulai dari memasukkan lembar jawaban, mendeteksi pilihan, melakukan koreksi, hingga menampilkan hasil dapat dilakukan secara terintegrasi. Pada bagian awal halaman terdapat judul “Koreksi Jawaban OMR” yang memberikan informasi mengenai fungsi utama aplikasi. Di bawah judul terdapat keterangan bahwa sistem digunakan untuk melakukan koreksi otomatis terhadap lembar jawaban 60 soal menggunakan teknologi Optical Mark Recognition. Selanjutnya tersedia bagian Upload Lembar Jawaban yang memungkinkan pengguna memasukkan citra lembar jawab melalui mekanisme drag and drop atau memilih file menggunakan tombol Browse files. Setelah citra berhasil dimasukkan, sistem menampilkan nama file yang digunakan sebagai data masukan.

Tahap berikutnya adalah bagian Masukkan Kunci Jawaban 60 Soal. Pada bagian ini pengguna dapat memasukkan kunci jawaban untuk 60 soal dengan lima alternatif pilihan, yaitu A, B, C, D, dan E. Kunci jawaban tersebut kemudian digunakan sebagai acuan pada proses koreksi. Setelah citra diproses, sistem memberikan informasi bahwa dua tabel lembar jawaban berhasil ditemukan. Sistem kemudian menampilkan Tabel Kiri untuk soal nomor 1–30 dan Tabel Kanan untuk soal nomor 31–60. Pembagian tersebut menunjukkan bahwa sistem mampu mengenali struktur lembar jawaban yang terdiri atas dua bagian.

Pada bagian Hasil Koreksi, sistem menampilkan ringkasan hasil pemeriksaan dalam bentuk beberapa informasi utama, yaitu jumlah jawaban benar, jumlah jawaban salah, jumlah jawaban kosong, dan nilai. Berdasarkan tampilan pada gambar, sistem menampilkan 8 jawaban benar, 52 jawaban salah, 0 jawaban kosong, serta nilai 13,33. Informasi tersebut memberikan hasil koreksi secara ringkas sehingga pengguna dapat mengetahui hasil ujian tanpa melakukan perhitungan secara manual. Sistem juga memberikan indikator bahwa seluruh soal telah terdeteksi memiliki jawaban.

Bagian selanjutnya adalah Hasil Koreksi pada Lembar Jawaban. Pada bagian ini sistem menampilkan kembali tabel jawaban dalam bentuk visual dan memberikan penanda pada posisi pilihan yang terdeteksi. Penanda berwarna hijau digunakan untuk menunjukkan pilihan jawaban yang terdeteksi sebagai jawaban peserta, sedangkan penanda merah digunakan sebagai indikator hasil yang tidak sesuai dengan kunci jawaban. Visualisasi ini menjadi bagian penting karena pengguna tidak hanya memperoleh nilai akhir, tetapi juga dapat melihat posisi jawaban yang diproses oleh sistem pada lembar jawaban.

Di bagian paling bawah terdapat Detail Jawaban yang menyajikan hasil pemeriksaan secara lebih rinci dalam bentuk tabel. Informasi yang ditampilkan meliputi nomor soal, jawaban yang terdeteksi, kunci jawaban, status, skor, dan margin. Status jawaban dibedakan menjadi BENAR dan SALAH. Dengan adanya tabel tersebut, proses koreksi dapat ditelusuri sampai pada tingkat setiap nomor soal. Pengguna dapat mengetahui jawaban yang diberikan peserta, membandingkannya dengan kunci jawaban, serta melihat hasil klasifikasi sistem.

Hasil implementasi menunjukkan bahwa sistem dapat memproses seluruh 50 Lembar Jawab pada dataset pengujian. Setiap Lembar Jawab menghasilkan 60 keputusan jawaban dan 300 area pilihan, sehingga keseluruhan proses menghasilkan 3.000 keputusan jawaban dan 15.000 area pilihan yang dianalisis. Seluruh 50 Lembar Jawab berhasil diproses tanpa kegagalan pada tahap validasi.

Meskipun demikian, keberhasilan pemrosesan seluruh dataset tidak secara langsung menunjukkan bahwa sistem telah robust terhadap berbagai kondisi pengambilan citra. Evaluasi pada penelitian ini terutama menggambarkan kinerja sistem pada karakteristik Lembar Jawab dan kondisi citra yang terdapat dalam dataset pengujian. Pengujian ketahanan terhadap variasi pencahayaan, rotasi ekstrem, jenis alat tulis, dan bayangan perlu dianalisis secara terpisah untuk mengetahui batas kemampuan pendekatan template-based yang digunakan.

3.2 Hasil Validasi Dataset

Pengujian dilakukan menggunakan 50 citra Lembar Jawab, dengan masing-masing Lembar Jawab terdiri atas 60 soal dan lima alternatif pilihan jawaban. Dengan demikian, terdapat 3.000 keputusan jawaban yang dibandingkan antara hasil identifikasi sistem dan ground truth. Selain keputusan jawaban, setiap Lembar Jawab memiliki 300 area pilihan yang dianalisis, sehingga keseluruhan dataset mencakup 15.000 area pilihan.

Hasil validasi menunjukkan bahwa seluruh 50 Lembar Jawab berhasil diproses oleh sistem. Tidak terdapat dataset yang mengalami kegagalan proses, sehingga tingkat keberhasilan pemrosesan Lembar Jawab mencapai 100%. Selanjutnya, hasil identifikasi pada 3.000 keputusan jawaban dibandingkan dengan ground truth untuk mengetahui tingkat ketepatan sistem.

Berdasarkan hasil evaluasi, sistem menghasilkan 2.863 prediksi yang sesuai dengan ground truth dan 137 prediksi yang tidak sesuai. Dengan demikian, accuracy yang diperoleh adalah sebesar 95,43%. Hasil evaluasi keseluruhan ditunjukkan pada Tabel 1.

Tabel 1. Hasil validasi dataset

Parameter	Hasil
Jumlah Lembar Jawab	50
Jumlah soal per Lembar Jawab	60
Total keputusan jawaban	3.000
Total area pilihan	15.000
Dataset berhasil diproses	50
Dataset gagal diproses	0



Parameter	Hasil
Prediksi benar	2.863
Prediksi salah	137
Accuracy	95,43%
Macro Precision	95,62%
Macro Recall	94,53%
Macro F1-score	94,92%

Accuracy sebesar 95,43% menunjukkan bahwa sebagian besar keputusan jawaban berhasil diidentifikasi sesuai dengan ground truth. Dari 3.000 keputusan yang dievaluasi, sebanyak 2.863 keputusan berhasil dikenali dengan benar, sedangkan 137 keputusan mengalami ketidaksesuaian. Dengan demikian, tingkat kesalahan klasifikasi pada keseluruhan dataset adalah sebesar 4,57%. Selain accuracy, evaluasi menggunakan macro precision, macro recall, dan macro F1-score. Sistem memperoleh macro precision sebesar 95,62%, macro recall sebesar 94,53%, dan macro F1-score sebesar 94,92%. Nilai macro recall yang sedikit lebih rendah dibandingkan macro precision menunjukkan bahwa masih terdapat beberapa data yang seharusnya masuk ke suatu kelas tetapi teridentifikasi sebagai kelas lain.

Penggunaan macro average dilakukan dengan memberikan bobot yang sama terhadap enam kelas keluaran, yaitu A, B, C, D, E, dan KOSONG. Oleh karena itu, nilai macro F1-score sebesar 94,92% tidak hanya merepresentasikan kemampuan sistem dalam mengenali pilihan A–E, tetapi juga mempertimbangkan kemampuan sistem dalam mengenali kondisi KOSONG.

Hasil tersebut menunjukkan bahwa pendekatan template-based yang diterapkan mampu menghasilkan kinerja klasifikasi yang cukup baik pada dataset pengujian. Namun, nilai accuracy dan macro F1-score tidak dapat digunakan sebagai satu-satunya dasar untuk menyimpulkan bahwa sistem telah robust terhadap seluruh kondisi pengambilan citra. Perbedaan karakteristik antar kelas dan pola kesalahan perlu dianalisis lebih lanjut, khususnya pada kelas KOSONG yang memiliki karakteristik berbeda dari pilihan jawaban A–E. Oleh karena itu, analisis berikutnya difokuskan pada performa masing-masing kelas untuk mengetahui distribusi kesalahan dan mengidentifikasi bagian algoritma yang masih menjadi sumber keterbatasan sistem.

3.3 Evaluasi Kinerja Klasifikasi

Evaluasi kinerja klasifikasi dilakukan untuk mengetahui kemampuan sistem dalam mengenali enam kelas keluaran, yaitu A, B, C, D, E, dan KOSONG. Evaluasi dilakukan menggunakan precision, recall, dan F1-score pada tingkat keputusan jawaban. Precision menunjukkan ketepatan sistem ketika memberikan suatu kelas, recall menunjukkan kemampuan sistem dalam menemukan data yang sebenarnya termasuk dalam suatu kelas, sedangkan F1-score menunjukkan keseimbangan antara precision dan recall. Hasil evaluasi kinerja untuk masing-masing kelas ditunjukkan pada Tabel 2.

Tabel 2. Hasil Evaluasi Kinerja Klasifikasi per Kelas

Kelas	Precision (%)	Recall (%)	F1-score (%)
A	93,12	94,75	93,93
B	98,60	90,97	94,63
C	95,36	97,81	96,57
D	95,62	97,94	96,77
E	91,04	100,00	95,31
KOSONG	100,00	85,71	92,31
Macro Average	95,62	94,53	94,92

Berdasarkan Tabel 2, kelas D memperoleh F1-score tertinggi sebesar 96,77%, dengan precision sebesar 95,62% dan recall sebesar 97,94%. Kelas C juga menunjukkan kinerja yang tinggi dengan F1-score sebesar 96,57%, precision sebesar 95,36%, dan recall sebesar 97,81%. Nilai recall yang tinggi pada kedua kelas menunjukkan bahwa sebagian besar jawaban yang sebenarnya termasuk kelas C dan D berhasil dikenali oleh sistem. Kelas A memperoleh precision sebesar 93,12%, recall sebesar 94,75%, dan F1-score sebesar 93,93%. Hasil tersebut menunjukkan bahwa kemampuan sistem dalam mengenali kelas A relatif seimbang antara ketepatan prediksi dan kemampuan menemukan data yang sebenarnya termasuk kelas A.

Kelas B memperoleh precision sebesar 98,60%, yang merupakan nilai precision tertinggi di antara kelas A–E. Namun, recall kelas B sebesar 90,97% menunjukkan bahwa masih terdapat sebagian jawaban yang sebenarnya termasuk kelas B tetapi teridentifikasi sebagai kelas lain. Dengan demikian, kesalahan pada kelas B lebih berkaitan dengan data kelas B yang tidak berhasil dipertahankan sebagai kelas B pada proses klasifikasi.

Kelas E memperoleh recall sebesar 100%, yang menunjukkan bahwa seluruh data yang termasuk kelas E pada ground truth berhasil dikenali sebagai E. Namun, precision kelas E sebesar 91,04% menunjukkan bahwa terdapat beberapa prediksi E yang sebenarnya berasal dari kelas lain. Kondisi tersebut menunjukkan adanya kecenderungan sebagian area dari kelas lain memperoleh nilai respons yang lebih tinggi pada kandidat E.

Kelas KOSONG memperoleh precision sebesar 100%, recall sebesar 85,71%, dan F1-score sebesar 92,31%. Nilai precision sebesar 100% menunjukkan bahwa setiap prediksi yang diberikan sistem sebagai KOSONG sesuai dengan ground truth. Namun, recall sebesar 85,71% menunjukkan bahwa sistem belum berhasil menemukan seluruh kondisi yang



sebenarnya kosong. Berdasarkan hasil evaluasi, kelas KOSONG memiliki 108 true positive, 0 false positive, dan 18 false negative.

Nilai false negative pada kelas KOSONG menunjukkan adanya kondisi ketika jawaban sebenarnya tidak memiliki tanda, tetapi sistem memberikan salah satu label A–E. Kondisi tersebut dapat terjadi ketika area kosong menghasilkan nilai respons yang cukup tinggi pada proses perbandingan dengan template. Nilai respons tersebut dapat dipengaruhi oleh keberadaan elemen cetak pada Lembar Jawab, seperti garis tabel dan karakter pilihan, maupun oleh perbedaan lokal antara citra Lembar Jawab dan template.

Selain itu, proses penyesuaian citra dengan template tidak selalu menghilangkan seluruh perbedaan posisi pada tingkat lokal. Pergeseran kecil pada garis tabel atau elemen cetak dapat menghasilkan perubahan piksel pada area yang dibandingkan. Apabila perubahan tersebut menghasilkan nilai respons yang melewati mathematical threshold, sistem dapat menetapkan area tersebut sebagai salah satu pilihan A–E meskipun secara ground truth berada pada kondisi KOSONG.

Secara keseluruhan, macro precision sebesar 95,62%, macro recall sebesar 94,53%, dan macro F1-score sebesar 94,92% menunjukkan bahwa sistem memiliki kinerja klasifikasi yang relatif baik pada enam kelas yang diuji. Namun, hasil per kelas menunjukkan bahwa karakteristik kesalahan tidak sama pada setiap kelas. Oleh karena itu, evaluasi lebih lanjut terhadap kelas KOSONG diperlukan untuk mengidentifikasi sumber kesalahan yang berkaitan dengan artefak visual, ketidaksesuaian lokal antara citra dan template, serta batas keputusan pada proses identifikasi.

3.4 Analisis Kelas KOSONG

Kelas KOSONG merupakan salah satu aspek penting dalam evaluasi karena sistem tidak hanya dituntut untuk mengenali pilihan A–E, tetapi juga harus mampu menentukan kondisi ketika peserta tidak memberikan tanda pada salah satu pilihan. Berdasarkan hasil pengujian terhadap 50 Lembar Jawab, kelas KOSONG memperoleh precision sebesar 100%, recall sebesar 85,71%, dan F1-score sebesar 92,31%. Nilai tersebut diperoleh dari 108 true positive, 0 false positive, dan 18 false negative.

Nilai precision sebesar 100% menunjukkan bahwa seluruh prediksi yang diberikan sistem sebagai KOSONG sesuai dengan kondisi pada ground truth. Dengan kata lain, sistem tidak memberikan label KOSONG pada jawaban yang sebenarnya termasuk kelas A–E. Sebaliknya, recall sebesar 85,71% menunjukkan bahwa masih terdapat 18 kondisi yang sebenarnya KOSONG tetapi tidak berhasil dikenali sebagai KOSONG. Kondisi tersebut menjadi sumber utama kesalahan pada kelas KOSONG.

False negative pada kelas KOSONG terjadi ketika sistem memberikan salah satu label A, B, C, D, atau E terhadap area yang pada ground truth tidak memiliki jawaban. Pola ini menunjukkan bahwa tantangan utama sistem bukan terletak pada munculnya prediksi KOSONG yang keliru, tetapi pada kemampuan membedakan perubahan visual yang benar-benar berasal dari tanda jawaban dengan perubahan visual yang terdapat pada area kosong.

Pada pendekatan template-based, area kosong tidak sepenuhnya memiliki nilai visual yang identik dengan template. Meskipun keduanya berasal dari desain Lembar Jawab yang sama, citra hasil pemotretan dapat memiliki perbedaan intensitas dan posisi pada tingkat piksel. Area kosong tetap mengandung elemen cetak seperti garis tabel, karakter pilihan, nomor soal, serta struktur geometris lainnya. Perbedaan antara elemen-elemen tersebut pada citra pengujian dan template dapat menghasilkan nilai perubahan ketika proses perbandingan dilakukan.

Selain perbedaan intensitas, ketidaktepatan penyesuaian pada tingkat lokal juga dapat memengaruhi hasil perbandingan. Proses ECC digunakan untuk menyesuaikan citra Lembar Jawab terhadap template, tetapi penyesuaian tersebut bekerja berdasarkan kesesuaian citra secara keseluruhan. Dengan demikian, masih dimungkinkan terdapat pergeseran lokal pada elemen tertentu, misalnya garis tabel atau batas area pilihan. Pergeseran tersebut dapat menghasilkan perubahan piksel pada area yang seharusnya dikategorikan sebagai KOSONG.

Perubahan visual yang dihasilkan oleh artefak cetak, perbedaan intensitas, noise, atau ketidaksesuaian lokal tersebut dapat meningkatkan nilai respons pada suatu area pilihan. Apabila nilai respons tertinggi melewati mathematical threshold, sistem akan mengategorikan area tersebut sebagai TERISI dan memilih kandidat dengan nilai respons tertinggi. Akibatnya, kondisi yang sebenarnya KOSONG dapat diklasifikasikan sebagai salah satu pilihan A–E dan menghasilkan false negative pada kelas KOSONG.

Temuan tersebut menunjukkan bahwa keberadaan template kosong belum sepenuhnya menjamin bahwa seluruh perubahan piksel yang terdeteksi merupakan tanda jawaban. Template berfungsi sebagai acuan untuk mengurangi pengaruh elemen visual yang bersifat tetap, tetapi efektivitasnya tetap dipengaruhi oleh tingkat kesesuaian antara citra pengujian dan template. Oleh karena itu, hasil recall kelas KOSONG sebesar 85,71% menunjukkan bahwa mekanisme perbandingan masih memiliki keterbatasan dalam memisahkan perubahan yang disebabkan oleh tanda jawaban dari perubahan yang berasal dari faktor visual lainnya.

Keterbatasan tersebut juga menunjukkan pentingnya penentuan mathematical threshold pada tahap identifikasi. Threshold yang terlalu rendah dapat menyebabkan perubahan kecil akibat noise, pergeseran garis, atau perbedaan pencahayaan dianggap sebagai tanda jawaban. Sebaliknya, threshold yang terlalu tinggi dapat menyebabkan tanda jawaban yang lemah tidak terdeteksi. Dengan demikian, nilai threshold menjadi salah satu faktor yang menentukan keseimbangan antara kemampuan mengenali jawaban terisi dan kemampuan mempertahankan kondisi KOSONG.

Berdasarkan hasil tersebut, peningkatan kinerja kelas KOSONG pada penelitian selanjutnya dapat diarahkan pada perbandingan yang lebih selektif terhadap perubahan piksel. Salah satu pendekatan yang dapat dipertimbangkan adalah menggunakan rasio perubahan piksel pada area jawaban, karakteristik tingkat kehitaman, atau distribusi perubahan di



dalam ROI sebagai tambahan informasi sebelum keputusan akhir diberikan. Pendekatan tersebut dapat membantu membedakan perubahan yang tersebar akibat artefak visual dengan perubahan yang terkonsentrasi pada area tanda jawaban.

Dengan demikian, hasil evaluasi kelas KOSONG menunjukkan bahwa pendekatan template-based telah mampu menghasilkan precision yang tinggi, tetapi masih terdapat kondisi kosong yang belum berhasil dikenali. Temuan ini menjadi dasar untuk mengevaluasi lebih lanjut pengaruh kondisi citra dan parameter keputusan terhadap stabilitas sistem, khususnya melalui analisis robustness dan keterbatasan yang dibahas pada subbab berikutnya.

3.5 Analisis Kesalahan Identifikasi

Analisis kesalahan identifikasi dilakukan untuk mengetahui karakteristik kesalahan yang terjadi pada proses klasifikasi jawaban. Berdasarkan hasil pengujian terhadap 3.000 keputusan jawaban, sistem menghasilkan 2.863 prediksi benar dan 137 prediksi salah. Dengan demikian, terdapat 137 keputusan yang tidak sesuai dengan ground truth atau sebesar 4,57% dari keseluruhan keputusan yang diuji.

Kesalahan identifikasi dapat terjadi ketika sistem memberikan kelas yang berbeda dari kelas sebenarnya. Pada pendekatan yang digunakan dalam penelitian ini, keputusan kelas ditentukan berdasarkan nilai respons pada masing-masing area pilihan setelah citra Lembar Jawab diselaraskan dengan template. Oleh karena itu, kesalahan klasifikasi dapat berkaitan dengan perbedaan nilai respons antararea pilihan.

Berdasarkan hasil evaluasi per kelas, pola kesalahan menunjukkan karakteristik yang berbeda pada setiap kelas. Kelas B memiliki precision sebesar 98,60% tetapi recall sebesar 90,97%. Kondisi tersebut menunjukkan bahwa sebagian besar prediksi B yang diberikan sistem memang benar, tetapi masih terdapat data yang sebenarnya merupakan kelas B namun diklasifikasikan sebagai kelas lain. Sebaliknya, kelas E memiliki recall sebesar 100%, tetapi precision hanya sebesar 91,04%. Hal tersebut menunjukkan bahwa seluruh data E pada ground truth berhasil ditemukan, tetapi terdapat beberapa prediksi E yang sebenarnya berasal dari kelas lain.

Kelas A memperoleh precision sebesar 93,12% dan recall sebesar 94,75%, sedangkan kelas C dan D masing-masing memiliki recall sebesar 97,81% dan 97,94%. Perbedaan antara precision dan recall pada setiap kelas menunjukkan bahwa kesalahan sistem tidak hanya berasal dari kegagalan menemukan suatu kelas, tetapi juga dapat terjadi karena sistem memberikan suatu label kepada data yang sebenarnya termasuk kelas lain.

Pada kelas KOSONG, pola kesalahan memiliki karakteristik yang berbeda. Precision sebesar 100% dan recall sebesar 85,71% menunjukkan bahwa sistem tidak menghasilkan false positive untuk kelas KOSONG, tetapi masih terdapat 18 false negative. Dengan demikian, kesalahan pada kelas KOSONG terutama terjadi ketika kondisi yang sebenarnya kosong memperoleh prediksi salah satu kelas A–E.

Pola tersebut dapat dikaitkan dengan mekanisme perbandingan template yang digunakan. Sistem menentukan respons berdasarkan perubahan visual antara citra Lembar Jawab dan template kosong. Perubahan yang muncul pada area jawaban tidak selalu disebabkan oleh tanda yang diberikan peserta. Perbedaan intensitas, noise, elemen cetak, atau ketidaksesuaian posisi lokal antara citra dan template dapat menghasilkan respons pada area tertentu. Jika respons tersebut lebih tinggi dibandingkan kandidat lainnya dan melewati mathematical threshold, sistem dapat memberikan label A–E terhadap area yang sebenarnya tidak memiliki jawaban.

Selain itu, perbedaan ketebalan atau karakteristik tanda jawaban dapat menyebabkan nilai respons antararea menjadi tidak seragam. Tanda yang lebih jelas dapat menghasilkan respons yang lebih tinggi, sedangkan tanda yang kurang kuat dapat menghasilkan respons yang mendekati batas keputusan. Kondisi tersebut dapat meningkatkan kemungkinan terjadinya kesalahan klasifikasi apabila perbedaan nilai antar kandidat relatif kecil.

3.6 Analisis Robustness dan Keterbatasan Sistem

Robustness merupakan kemampuan sistem untuk mempertahankan kinerja ketika terdapat perubahan kondisi pada citra masukan. Pada penelitian ini, evaluasi utama dilakukan menggunakan 50 Lembar Jawab yang menghasilkan 3.000 keputusan jawaban. Seluruh Lembar Jawab berhasil diproses dan sistem memperoleh accuracy sebesar 95,43%, macro precision sebesar 95,62%, macro recall sebesar 94,53%, dan macro F1-score sebesar 94,92%. Hasil tersebut menunjukkan kinerja sistem pada karakteristik dataset yang digunakan dalam penelitian.

Namun, hasil tersebut belum dapat digunakan untuk menyatakan bahwa sistem telah robust terhadap seluruh variasi kondisi pengambilan citra. Penelitian ini belum melakukan pengujian robustness secara terkontrol dengan memberikan variasi khusus pada tingkat pencahayaan, rotasi ekstrem, jenis alat tulis, dan keberadaan bayangan. Oleh karena itu, pengaruh masing-masing kondisi tersebut terhadap kinerja sistem belum dapat diukur secara kuantitatif pada penelitian ini.

Keterbatasan tersebut penting diperhatikan karena pendekatan yang digunakan bergantung pada perbandingan citra Lembar Jawab dengan template kosong. Perubahan kondisi pencahayaan dapat mengubah distribusi intensitas piksel antara citra pengujian dan template. Meskipun adaptive thresholding digunakan pada tahap preprocessing, perubahan intensitas pada tahap perbandingan template tetap berpotensi menghasilkan perbedaan visual pada area yang seharusnya tidak memiliki tanda jawaban.

Rotasi atau perubahan posisi Lembar Jawab juga dapat memengaruhi kesesuaian antara citra pengujian dan template. Koreksi perspektif dan penyelarasan menggunakan ECC digunakan untuk mengurangi perbedaan geometris tersebut, tetapi keduanya tidak menjamin bahwa seluruh elemen pada citra akan memiliki kesesuaian posisi secara lokal.



Pergeseran kecil pada garis tabel atau batas area pilihan dapat menghasilkan perubahan piksel yang selanjutnya memengaruhi nilai respons pada proses perbandingan.

Jenis alat tulis juga berpotensi memengaruhi nilai respons area jawaban. Pensil dan pulpen dapat menghasilkan karakteristik intensitas, ketebalan, dan distribusi tanda yang berbeda. Perbedaan tersebut dapat menyebabkan nilai respons pada area yang terisi menjadi tidak seragam. Namun, karena penelitian ini belum melakukan pengujian terkontrol berdasarkan jenis alat tulis, pengaruh faktor tersebut terhadap accuracy atau recall belum dapat disimpulkan secara empiris.

Keberadaan bayangan pada citra juga merupakan faktor yang berpotensi memengaruhi proses perbandingan. Bayangan dapat menyebabkan perubahan intensitas pada sejumlah area citra secara bersamaan dan menghasilkan perbedaan terhadap template kosong. Kondisi tersebut dapat meningkatkan nilai respons pada area tertentu meskipun tidak terdapat tanda jawaban. Pengaruh bayangan menjadi semakin relevan pada analisis kelas KOSONG karena sistem harus membedakan perubahan visual akibat kondisi pencahayaan dari perubahan yang disebabkan oleh tanda jawaban.

Berdasarkan hasil evaluasi, keterbatasan yang telah teridentifikasi secara langsung adalah masih adanya 18 false negative pada kelas KOSONG. Kondisi tersebut menunjukkan bahwa sebagian jawaban yang sebenarnya kosong masih menghasilkan prediksi A–E. Temuan ini menunjukkan bahwa perbandingan terhadap template belum sepenuhnya mampu memisahkan perubahan yang berkaitan dengan tanda jawaban dari perubahan visual lainnya.

Dengan demikian, hasil penelitian ini lebih tepat diinterpretasikan sebagai evaluasi kinerja sistem pada dataset yang digunakan, bukan sebagai bukti bahwa metode telah robust terhadap seluruh kondisi lingkungan. Pengujian robustness yang lebih terkontrol diperlukan pada penelitian selanjutnya dengan membuat kelompok pengujian berdasarkan variasi pencahayaan, tingkat rotasi, jenis alat tulis, dan kondisi bayangan. Setiap kelompok dapat dibandingkan menggunakan accuracy, precision, recall, dan F1-score, dengan perhatian khusus pada recall kelas KOSONG.

Selain pengujian kondisi citra, pengembangan selanjutnya dapat mengevaluasi pengaruh mathematical threshold terhadap hasil klasifikasi. Pengujian beberapa nilai threshold dapat digunakan untuk mengetahui hubungan antara kemampuan sistem dalam mempertahankan jawaban KOSONG dan kemampuannya dalam mengenali jawaban A–E. Pendekatan tersebut dapat menjadi dasar untuk memperoleh nilai threshold yang lebih sesuai dengan karakteristik perubahan piksel pada Lembar Jawab.

Dengan pendekatan tersebut, keterbatasan robustness pada penelitian ini tidak mengurangi hasil evaluasi yang telah diperoleh, tetapi menjadi batas interpretasi terhadap kinerja sistem. Hasil penelitian menunjukkan bahwa pendekatan template-based dapat digunakan untuk melakukan koreksi otomatis pada dataset yang diuji, sementara pengujian pada kondisi citra yang lebih beragam masih diperlukan untuk memastikan kestabilan sistem pada penggunaan yang lebih luas.

3.7 Pembahasan

Hasil penelitian menunjukkan bahwa pendekatan Optical Mark Recognition (OMR) berbasis template-based dapat digunakan untuk melakukan identifikasi dan koreksi otomatis pada Lembar Jawab pilihan ganda. Pengujian terhadap 50 Lembar Jawab menghasilkan 3.000 keputusan jawaban, dengan 2.863 keputusan berhasil diidentifikasi sesuai dengan ground truth dan 137 keputusan tidak sesuai. Sistem memperoleh accuracy sebesar 95,43%, macro precision sebesar 95,62%, macro recall sebesar 94,53%, dan macro F1-score sebesar 94,92%. Hasil tersebut menunjukkan bahwa pendekatan yang digunakan mampu memberikan kinerja klasifikasi yang relatif baik pada dataset pengujian.

Kinerja tersebut diperoleh melalui rangkaian proses pengolahan citra yang mencakup preprocessing, grayscale, adaptive thresholding, deteksi kontur, koreksi perspektif, segmentasi area jawaban, penyelarasan menggunakan ECC, perbandingan dengan template kosong, dan penentuan jawaban berdasarkan mathematical threshold. Penggunaan template kosong memberikan acuan visual terhadap kondisi Lembar Jawab sebelum diisi. Dengan demikian, proses identifikasi tidak hanya bergantung pada tingkat intensitas absolut dari suatu area, tetapi mempertimbangkan perubahan visual antara citra pengujian dan citra acuan.

Penyelarasan menggunakan ECC memiliki peran penting dalam pendekatan tersebut karena posisi citra Lembar Jawab hasil akuisisi tidak selalu identik dengan posisi template. Koreksi perspektif terlebih dahulu digunakan untuk mengurangi perubahan geometris secara umum, sedangkan penyelarasan dengan ECC digunakan untuk meningkatkan kesesuaian antara citra dan template sebelum proses perbandingan. Meskipun demikian, hasil evaluasi kelas KOSONG menunjukkan bahwa penyelarasan global belum sepenuhnya menghilangkan perbedaan lokal pada area jawaban. Oleh karena itu, keberadaan ECC membantu mengurangi ketidaksesuaian posisi, tetapi tidak dapat dianggap sebagai mekanisme yang sepenuhnya menghilangkan seluruh artefak visual.

Hasil evaluasi per kelas memperlihatkan bahwa kinerja sistem relatif konsisten pada pilihan A–E. Kelas C dan D memperoleh recall masing-masing sebesar 97,81% dan 97,94%, sedangkan kelas E memperoleh recall sebesar 100%. Kelas B memiliki precision sebesar 98,60%, tetapi recall sebesar 90,97%, sedangkan kelas A memperoleh precision sebesar 93,12% dan recall sebesar 94,75%. Perbedaan tersebut menunjukkan bahwa meskipun sistem mampu mengenali sebagian besar pilihan jawaban, masih terdapat ketidakseimbangan antara kemampuan menemukan data suatu kelas dan ketepatan ketika memberikan label kelas tersebut.

Temuan yang paling penting terdapat pada kelas KOSONG. Kelas tersebut memperoleh precision sebesar 100%, recall sebesar 85,71%, dan F1-score sebesar 92,31%. Terdapat 108 true positive, 0 false positive, dan 18 false negative. Pola tersebut menunjukkan bahwa sistem sangat selektif ketika memberikan label KOSONG, tetapi belum mampu menemukan seluruh kondisi yang sebenarnya kosong. Dengan kata lain, kesalahan lebih banyak terjadi ketika area yang



seharusnya KOSONG memperoleh respons yang cukup tinggi dan kemudian diklasifikasikan sebagai salah satu pilihan A–E.

Kondisi tersebut berkaitan dengan karakteristik pendekatan template-based. Template kosong dapat digunakan untuk merepresentasikan elemen visual yang bersifat tetap, tetapi citra aktual masih dapat memiliki perbedaan intensitas, noise, ketidaktepatan posisi, dan perubahan lokal. Garis tabel dan karakter cetak yang mengalami sedikit pergeseran antara citra pengujian dan template dapat menghasilkan perubahan piksel pada proses perbandingan. Perubahan tersebut berpotensi meningkatkan nilai respons pada area yang sebenarnya tidak memiliki tanda jawaban.

Mathematical threshold menjadi bagian penting dalam kondisi tersebut. Sistem harus menentukan batas antara perubahan visual yang dianggap sebagai tanda jawaban dan perubahan yang dianggap sebagai variasi citra. Threshold yang terlalu rendah dapat meningkatkan kemungkinan perubahan kecil akibat noise atau ketidaksesuaian template dianggap sebagai tanda. Sebaliknya, threshold yang terlalu tinggi dapat menyebabkan tanda jawaban dengan intensitas rendah tidak terdeteksi. Oleh karena itu, pemilihan threshold berpengaruh langsung terhadap keseimbangan antara pengenalan kelas A–E dan kemampuan mempertahankan kelas KOSONG. Hasil penelitian juga menunjukkan bahwa accuracy sebesar 95,43% belum cukup untuk menyatakan bahwa sistem memiliki ketahanan terhadap seluruh kondisi penggunaan. Evaluasi yang dilakukan menunjukkan kinerja pada dataset yang digunakan, tetapi belum mencakup pengujian robustness secara terkontrol terhadap variasi pencahayaan, rotasi ekstrem, jenis alat tulis, dan bayangan. Faktor-faktor tersebut berpotensi mengubah karakteristik piksel yang menjadi dasar perbandingan dengan template. Oleh sebab itu, generalisasi hasil penelitian perlu dibatasi pada karakteristik dataset dan kondisi citra yang diuji.

Secara keseluruhan, hasil penelitian menunjukkan bahwa pendekatan template-based memberikan alternatif yang dapat digunakan untuk implementasi OMR tanpa memerlukan model pembelajaran mesin. Pendekatan ini memiliki keuntungan berupa proses identifikasi yang didasarkan pada struktur Lembar Jawab dan perbandingan langsung dengan template. Namun, ketergantungan terhadap kesesuaian visual antara citra pengujian dan template menjadi salah satu keterbatasan ketika terdapat perubahan lokal pada citra. Berdasarkan temuan tersebut, pengembangan selanjutnya dapat diarahkan pada penyempurnaan mekanisme diferensiasi antara tanda jawaban dan artefak visual. Perhitungan rasio perubahan piksel, tingkat kehitaman area, distribusi perubahan dalam ROI, serta optimasi mathematical threshold dapat.

4. KESIMPULAN

Berdasarkan hasil implementasi dan evaluasi, sistem Optical Mark Recognition (OMR) berbasis Computer Vision dengan pendekatan template-based mampu melakukan identifikasi dan koreksi otomatis terhadap Lembar Jawab pilihan ganda. Sistem memproses citra melalui tahapan preprocessing, grayscale, adaptive thresholding, deteksi kontur, koreksi perspektif, segmentasi area jawaban, penyalarsan menggunakan Enhanced Correlation Coefficient (ECC), perbandingan dengan template kosong, penentuan mathematical threshold, serta koreksi berdasarkan kunci jawaban. Pengujian terhadap 50 Lembar Jawab dengan 60 soal per lembar menghasilkan 3.000 keputusan jawaban dan 15.000 area pilihan yang berhasil diproses tanpa kegagalan dataset. Hasil evaluasi menunjukkan bahwa sistem memperoleh accuracy sebesar 95,43%, macro precision sebesar 95,62%, macro recall sebesar 94,53%, dan macro F1-score sebesar 94,92%. Hasil tersebut menunjukkan bahwa pendekatan template-based dapat digunakan untuk melakukan identifikasi pilihan A, B, C, D, E, dan KOSONG pada struktur Lembar Jawab yang digunakan dalam penelitian. Analisis per kelas menunjukkan bahwa keterbatasan utama sistem terdapat pada kemampuan mempertahankan klasifikasi KOSONG. Kelas KOSONG memperoleh precision sebesar 100%, recall sebesar 85,71%, dan F1-score sebesar 92,31%, dengan 108 true positive, 0 false positive, dan 18 false negative. Kondisi tersebut menunjukkan bahwa sistem tidak memberikan label KOSONG secara keliru terhadap jawaban yang sebenarnya terisi, tetapi masih terdapat kondisi kosong yang teridentifikasi sebagai salah satu pilihan A–E. Secara kausal, kesalahan tersebut tidak dapat dikaitkan hanya dengan kegagalan ECC. ECC berfungsi untuk meningkatkan kesesuaian posisi citra Lembar Jawab terhadap template berdasarkan karakteristik visual secara keseluruhan. Namun, kesesuaian secara global tidak menjamin bahwa seluruh elemen pada tingkat lokal memiliki posisi dan intensitas yang identik. Pergeseran kecil pada garis tabel, perbedaan intensitas akibat proses akuisisi citra, noise, serta karakteristik elemen cetak dapat menghasilkan perubahan piksel pada area yang sebenarnya kosong. Perubahan tersebut kemudian dapat menghasilkan nilai respons yang cukup tinggi pada proses perbandingan template. Apabila nilai tersebut melewati mathematical threshold, area yang sebenarnya KOSONG dapat diklasifikasikan sebagai salah satu pilihan A–E. Dengan demikian, sumber keterbatasan sistem lebih tepat dipahami sebagai interaksi antara ketidaksempurnaan penyalarsan lokal, perbedaan karakteristik citra terhadap template, dan mekanisme threshold pada tahap pengambilan keputusan. ECC membantu mengurangi perbedaan geometris, tetapi tidak secara langsung membedakan apakah perubahan piksel berasal dari tanda jawaban atau artefak visual. Hal tersebut menjelaskan mengapa sistem tetap dapat menghasilkan false negative pada kelas KOSONG meskipun proses alignment telah diterapkan. Penelitian ini juga belum melakukan robustness test secara terkontrol terhadap variasi pencahayaan, rotasi ekstrem, jenis alat tulis, dan bayangan. Oleh karena itu, accuracy sebesar 95,43% harus dipahami sebagai kinerja sistem pada karakteristik dataset yang diuji dan belum dapat digunakan untuk menyatakan bahwa sistem robust terhadap seluruh kondisi pengambilan citra. Secara keseluruhan, penelitian membuktikan bahwa pendekatan template-based dapat digunakan sebagai metode OMR berbasis Computer Vision untuk melakukan koreksi otomatis pada format Lembar Jawab



yang diuji. Namun, peningkatan reliabilitas sistem terutama pada deteksi KOSONG masih memerlukan penyempurnaan mekanisme diferensiasi antara tanda jawaban dan perubahan visual yang bukan berasal dari tanda peserta.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Kontribusi setiap penulis pada penelitian dan penulisan manuskrip ini dinyatakan sebagai berikut:

Konseptualisasi: N. T., O. D. S., dan A. S.;

Metodologi: N. T., O. D. S., dan A. S.;

Perangkat Lunak: N. T. dan O. D. S.;

Validasi: N. T., O. D. S., dan A. S.;

Analisis Formal: N. T. dan A. S.;

Investigasi: N. T., O. D. S., A. S., dan A. M.;

Sumber Daya: N. T., O. D. S., A. S., dan A. M.;

Kurasi Data: N. T. dan O. D. S.;

Penulisan Draf Awal: N. T.;

Penulisan, Peninjauan, dan Penyuntingan: N. T., O. D. S., A. S., dan A. M.;

Visualisasi: N. T. dan O. D. S.;

Supervisi: O. D. S. dan A. S.

Seluruh penulis bertanggung jawab atas kontribusi masing-masing serta telah meninjau dan menyetujui versi akhir manuskrip sebelum diajukan untuk publikasi.

5.2 Ketersediaan Data

Data yang digunakan dalam penelitian ini dikumpulkan dari responden yang berpartisipasi dalam penelitian dan digunakan untuk mendukung pengujian serta evaluasi sistem *Optical Mark Recognition* (OMR). Data yang mendukung temuan penelitian ini dapat diperoleh melalui permintaan yang wajar kepada penulis korespondensi dengan tetap memperhatikan kerahasiaan identitas responden dan ketentuan penggunaan data yang berlaku.

5.3 Pendanaan

Penelitian ini didanai melalui program Basis Informasi Penelitian dan Pengabdian kepada Masyarakat (BIMA) yang diselenggarakan oleh Kementerian Pendidikan Tinggi, Sains, dan Teknologi (Kemdiktisaintek), Republik Indonesia.

5.4 Pernyataan Dewan Peninjau Institusional

Penelitian ini berfokus pada implementasi dan evaluasi sistem Optical Mark Recognition (OMR) untuk koreksi otomatis lembar jawaban pilihan ganda. Data penelitian dikumpulkan dari responden dan digunakan untuk menguji kinerja sistem dalam mengenali pilihan jawaban. Informasi mengenai persetujuan atau pengecualian dari dewan peninjau institusional belum dilaporkan secara terpisah dalam naskah ini.

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Keikutsertaan responden dalam penelitian ini berkaitan dengan pengumpulan data yang digunakan untuk pengujian dan evaluasi sistem OMR. Data responden digunakan untuk mendukung analisis kinerja sistem dalam membaca lembar jawaban pilihan ganda. Informasi mengenai persetujuan berdasarkan informasi (informed consent) dari responden belum dilaporkan secara terpisah dalam naskah ini. Kerahasiaan identitas responden perlu dijaga dalam pengolahan, penyajian, dan publikasi data penelitian.

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Dalam proses penulisan dan revisi manuskrip, penulis menggunakan teknologi AI generatif sebagai alat bantu untuk meningkatkan kejelasan bahasa dan penyuntingan naskah. Konten yang dihasilkan dengan bantuan AI telah ditinjau dan disunting oleh penulis. AI tidak menggantikan peran penulis dalam pengambilan keputusan ilmiah, analisis data, maupun interpretasi hasil penelitian. Seluruh penulis bertanggung jawab penuh atas keakuratan isi, referensi, interpretasi hasil, dan versi akhir manuskrip yang diajukan untuk publikasi.

REFERENCES

- [1] S. Mondal, P. De, S. Malakar, and R. Sarkar, "OMRNet: A lightweight deep learning model for optical mark recognition," *Multimed. Tools Appl.*, vol. 83, no. 5, pp. 14011–14045, Jul. 2023, doi: 10.1007/s11042-023-15408-8.





- [2] Q. Hafeez, W. Aslam, M. Ikramullah Lali, S. Ahmad, M. Alqahtani, and M. Shafiq, "Fault Tolerant Optical Mark Recognition," *Computers, Materials & Continua*, vol. 74, no. 2, pp. 3829–3847, 2023, doi: 10.32604/cmc.2023.026422.
- [3] M. Ali, S. Hammoud, and A. Esper, "Optical Mark Recognition Techniques for Multiple-Choice Tests: A Comprehensive Review," *IEEE Access*, vol. 13, pp. 156407–156423, 2025, doi: 10.1109/ACCESS.2025.3600106.
- [4] Y. Sancar, "Detection of risk levels in optical mark recognition exams using categorical feature-based machine learning models," *Sci. Rep.*, vol. 15, no. 1, p. 44611, Dec. 2025, doi: 10.1038/s41598-025-32963-2.
- [5] Y. Hernández-Mier, M. A. Nuño-Maganda, S. Polanco-Martagón, G. Acosta-Villarreal, and R. Posada-Gómez, "Unsupervised Optical Mark Recognition on Answer Sheets for Massive Printed Multiple-Choice Tests," *J. Imaging*, vol. 11, no. 9, p. 308, Sep. 2025, doi: 10.3390/jimaging11090308.
- [6] R. Rasmuin, "Does the policy regarding the role of the national examination in determining graduation affect student achievement?," *Jurnal Penelitian dan Evaluasi Pendidikan*, vol. 26, no. 2, Dec. 2022, doi: 10.21831/pep.v26i2.54422.
- [7] E. R. Ali and N. M. Shati, "Evaluation System for Multiple-Choice Questions Using Optical Mark Recognition: A Survey," *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 17, no. 1, pp. 202–213, Mar. 2025, doi: 10.29304/jqscsm.2025.17.11975.
- [8] M. Yusuf, A. Ihsan Pradana, and R. Susanto, "Computer Vision-Based Optical Mark Recognition (OMR) System for Automated Exam Answer Sheet Correction," *Journal of Artificial Intelligence and Software Engineering*, vol. 5, no. 3, pp. 962–971, 2025, doi: 10.30811/jaise.v5i3.7142.
- [9] E. Osagie, W. Ji, and N. Helian, "Burnt-in Text Recognition from Medical Imaging Modalities : Existing Machine Learning Practices," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 28, no. 1, pp. 103–110, 2024, doi: 10.20965/jaciii.2024.p0103.
- [10] R. C. Dharmik, S. Rangari, S. Jain, A. Nilawar, G. Deshmukh, and A. Yeole, "Optical Mark Recognition Evaluation System using Dual-Component Approach," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 10s, pp. 349–353, 2024.
- [11] S. Mondal, P. De, S. Malakar, and R. Sarkar, "OMRNet: A lightweight deep learning model for optical mark recognition," *Multimed. Tools Appl.*, vol. 83, no. 5, pp. 14011–14045, 2024, doi: 10.1007/s11042-023-15408-8.
- [12] K. Tabassum and Z. Rahman, "Optical Mark Recognition with Object Detection and Clustering," in *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, IEEE, May 2024, pp. 352–357. doi: 10.1109/ICEEICT62016.2024.10534561.
- [13] Sunardi, A. Fadlil, and N. Trisanti, "Comparative analysis of euclidean, manhattan, canberra, and squared chord methods in face recognition," *Revue d'Intelligence Artificielle*, vol. 37, no. 3, pp. 593–599, 2023, doi: 10.18280/ria.370308.
- [14] Yaseen, O. J. Kwon, J. Kim, S. Jamil, J. Lee, and F. Ullah, "Next-Gen Dynamic Hand Gesture Recognition: MediaPipe, Inception-v3 and LSTM-Based Enhanced Deep Learning Model," *Electronics (Switzerland)*, vol. 13, no. 16, pp. 1–11, 2024, doi: 10.3390/electronics13163233.
- [15] M. N. Hermawan, "Deteksi Lembar Jawaban Komputer Menggunakan OMR (Optical Mark Recognition) Di MTS Nurul Iman," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 3, pp. 1361–1372, 2021, doi: 10.35957/jatisi.v8i3.1369.
- [16] F. Marpaung, F. Aulia, and R. C. Nabila, *Computer Vision Dan Pengolahan Citra Digital*. 2022.
- [17] F. E. Darmawan, S. K. Magfiroh, M. K. Hanifah, and M. Saman, "Menguji Keefektifan Metode Mask R-CNN, dan Metode Keypoint R-CNN dalam Deteksi Objek Citra.," *Jurnal Komputer Dan Teknologi Informasi*, vol. 2, no. 2, pp. 84–93, 2024, doi: 10.26714/jkti.v2i2.14084.
- [18] S. Bao, H. Liu, L. Min, D. Xu, and G. Le, "Color-to-Gray Conversion Method Based on Chroma Difference," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 28, no. 3, pp. 655–667, May 2024, doi: 10.20965/jaciii.2024.p0655.
- [19] T. Wu, C. Eising, M. Glavin, and E. Jones, "An Efficient and Effective Image Decolorization Algorithm Based on Cumulative Distribution Function," *J. Imaging*, vol. 10, no. 3, p. 51, Feb. 2024, doi: 10.3390/jimaging10030051.
- [20] N. Detsikas, N. Mitianoudis, and N. Papamarkos, "A Dilated MultiRes Visual Attention U-Net for historical document image binarization," *Signal Process. Image Commun.*, vol. 122, p. 117102, Mar. 2024, doi: 10.1016/j.image.2024.117102.
- [21] R. Ahila Priyadarshini and H. Shriram, "Historical Document Image Binarization Using Support Vector Machine and Adaptive Threshold Selection," 2024, pp. 361–374. doi: 10.1007/978-981-97-5866-1_26.
- [22] S. K. Wen Hwooi and A. Q. Md Sabri, "INVESTIGATION OF IMAGE STITCHING REFINEMENT WITH ENHANCED CORRELATION COEFFICIENT," *Malaysian Journal of Computer Science*, vol. 33, no. 1, pp. 22–34, Jan. 2020, doi: 10.22452/mjcs.vol33no1.2.
- [23] S. Szabó, I. J. Holb, V. É. Abriha-Molnár, G. Szatmári, S. K. Singh, and D. Abriha, "Classification Assessment Tool: A program to measure the uncertainty of classification models in terms of class-level metrics," *Appl. Soft Comput.*, vol. 155, p. 111468, Apr. 2024, doi: 10.1016/j.asoc.2024.111468.
- [24] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, Mar. 2024, doi: 10.1038/s41598-024-56706-x.