

BAB I

PENDAHULUAN

1.1 Latar Belakang

Manuskrip lontar adalah warisan budaya yang harus dijaga agar tidak hilang oleh perkembangan zaman. Karya sastra di dalamnya mencerminkan kehidupan masyarakat. Lontar juga mengandung berbagai karya sastra yang memberikan pelajaran tentang filosofi hidup, agama, serta berbagai ilmu lain yang sangat berguna untuk kehidupan (Widiarti & Suparwito., 2022). Oleh karena itu, penting untuk melestarikan manuskrip lontar agar generasi mendatang dapat terus belajar dan memahami nilai-nilai budaya yang terkandung di dalamnya. Jika tidak dilakukan upaya pelestarian, manuskrip ini bisa terancam punah, baik akibat kerusakan fisik maupun berkurangnya minat masyarakat terhadap tradisi lokal. Pelestarian tidak hanya menjaga fisik manuskrip, tetapi juga memanfaatkan teknologi modern, seperti digitalisasi, untuk mendokumentasikan dan memudahkan akses terhadapnya.

Pelestarian naskah lontar Bali semakin menjadi perhatian seiring dengan semakin banyaknya generasi muda yang tidak mampu membacanya. Ribuan naskah lontar warisan leluhur Bali yang mengandung nilai-nilai adiluhung nyatanya belum banyak diminati oleh generasi muda, baik dari segi kemampuan menulis, membaca, maupun mengkaji isinya (Widiantana et al., 2022). Kondisi ini sejalan dengan temuan yang mencatat bahwa sebanyak 52,86% orang tua siswa menyatakan materi aksara Bali yang menjadi kunci untuk membaca lontar dirasakan sulit oleh anak-

anak mereka (Putu Listia et al., 2022). Minat yang terbatas ini berdampak langsung pada kondisi fisik lontar itu sendiri; Data penyuluh bahasa Bali Provinsi Bali menunjukkan bahwa dari 8.366 lontar yang tercatat di seluruh Bali, sebanyak 2.562 lontar telah dalam kondisi rusak, karena lama tidak dibaca dan tidak dirawat (Ardiyasa, 2021). Tidak hanya itu, generasi Bali tidak hanya meninggalkan naskah lontar sebagai benda peninggalan, tetapi juga semakin melupakan aksara yang digunakan untuk membaca dan memahaminya, bahkan para pemilik lontar secara pribadi kerap tidak lagi mampu merawat maupun membaca naskah yang mereka miliki (Antara et.al., 2012). Tantangan dalam pelestarian lontar tidak hanya bersumber dari aspek fisik naskah yang semakin menurun kualitasnya, tetapi juga dari tradisi yang memandang pembacaan aksara Bali sebagai sesuatu yang sakral dan tabu untuk dilakukan (Kesiman & Dermawan, 2021).

Kendati Bahasa Bali masih digunakan secara aktif dalam percakapan sehari-hari, penggunaan aksara Bali kian tidak meluas. Kondisi ini membuat kemampuan generasi muda dalam menulis dan membaca aksara Bali menurun secara bertahap, sehingga filolog yang mahir membaca aksara Bali pada koleksi lontar menjadi semakin sulit ditemukan (Kesiman & Dermawan, 2021). Minimnya jumlah ahli filologi yang mampu membaca naskah beraksara Bali, ditambah rendahnya pemahaman generasi masa kini terhadap aksara tersebut, mengakibatkan koleksi lontar di museum hanya tersimpan begitu saja setelah melalui proses katalogisasi. Di sisi lain, kekhawatiran terhadap kerusakan fisik lontar turut memperlambat upaya penyebaran pengetahuan yang terkandung di dalamnya, padahal isi lontar tersebut berpotensi memberikan manfaat besar bagi kehidupan masa kini.

Upaya pelestarian warisan budaya tulis seperti manuskrip lontar aksara Bali menjadi semakin penting seiring dengan perkembangan teknologi. Salah satu langkah yang dapat diambil adalah dengan mengembangkan model machine learning untuk mengklasifikasikan lontar. Teknologi pembelajaran mesin memberikan peluang besar dalam hal ini, termasuk penggunaan *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)* sebagai metode klasifikasi. Dengan kedua metode tersebut, aksara Bali pada manuskrip lontar dapat dikenali dan diklasifikasikan secara otomatis, sehingga membantu proses digitalisasi dan pelestarian naskah-naskah kuno ini.

Dalam penelitian ini, akan dilakukan pengujian terhadap dua pendekatan, yaitu klasifikasi langsung menggunakan CNN secara mandiri, dan pendekatan hybrid CNN-SVM di mana CNN digunakan sebagai ekstraktor fitur dan SVM sebagai algoritma klasifikasi. Pengujian terhadap CNN saja dilakukan untuk menilai sejauh mana model *deep learning* tersebut mampu melakukan klasifikasi aksara Bali secara end-to-end, serta menjadi baseline untuk dibandingkan dengan performa model gabungan.

Model klasifikasi aksara Bali pada manuskrip lontar ini nantinya akan ditujukan untuk mendukung para peneliti filologi, serta pengembang aplikasi pelestarian budaya dalam melakukan digitalisasi, klasifikasi, dan pencarian informasi berbasis aksara Bali. Pihak yang dapat memvalidasi hasil klasifikasi model ini mencakup ahli bahasa dan budaya Bali, pakar filologi, serta pengelola museum dan perpustakaan yang memiliki pengetahuan mendalam tentang aksara Bali dan struktur naskah lontar.

Penelitian yang relevan adalah kompetisi *Document Image Analysis Tasks for Southeast Asian Palm Leaf Manuscripts* yang diselenggarakan dalam konteks ICFHR 2018 (Kesiman et al., 2018). Penelitian ini menggunakan dataset AMADI LontarSet dengan 133 kelas karakter, 11.710 citra latih, dan 7.673 citra uji aksara Bali terisolasi pada naskah lontar. Metode terbaik pada kompetisi ini diajukan oleh Group 20 (G20) yang menggunakan jaringan saraf tiruan konvolusional sangat dalam (*very deep CNN*) sebanyak 100 lapisan dengan koneksi padat (*dense connection*) serta strategi augmentasi data untuk mengatasi keterbatasan data dan kemiripan antar karakter. Hasil evaluasi menunjukkan bahwa metode G20 berhasil mencapai tingkat pengenalan tertinggi sebesar 92,17% pada dataset aksara Bali, melampaui metode terbaik dari kompetisi ICFHR 2016 yang hanya mencapai 88,39%, serta metode berbasis CNN dan fitur buatan tangan (*handcrafted features*) yang masing-masing hanya mencapai 85,39% dan 85,63%. Capaian akurasi ini menjadikan metode G20 sebagai acuan tertinggi (*state-of-the-art*) pada dataset yang sama hingga saat ini.

Penelitian lain yang relevan adalah studi tentang pengenalan karakter aksara Bali terisolasi pada naskah lontar menggunakan model CNN VGG16, MobileNetV1, dan Simple CNN yang dilatih dari awal (*training from scratch*). Penelitian ini menggunakan dataset *Isolated Character Recognition of Balinese Script in Palm Leaf Manuscript Images* (Challenge-3-ForTrain.zip) yang terdiri dari 133 karakter dengan 10.540 data rekaman sebagai data latih. Hasil penelitian menunjukkan bahwa model VGG16 mampu mencapai akurasi pelatihan tertinggi sebesar 91,01%, akurasi pengujian 85,3%, dan akurasi validasi 83,93% hanya dalam 6 *epoch*. Meskipun demikian, penelitian ini memiliki keterbatasan berupa

perbedaan spesifikasi perangkat yang digunakan pada beberapa model sehingga tidak dapat membandingkan waktu yang dibutuhkan selama proses pelatihan, serta belum memproses hasil pengenalan menjadi rangkaian karakter yang membentuk kata atau kalimat utuh (Imawati et al., 2024).

Berdasarkan permasalahan terkait dengan penerapan metode modern dalam klasifikasi aksara Bali pada manuskrip lontar, penelitian ini, akan dilakukan pengujian terhadap dua pendekatan, yaitu klasifikasi langsung menggunakan CNN secara mandiri, dan pendekatan hybrid CNN–SVM. Pengujian terhadap CNN saja dilakukan untuk menilai sejauh mana model *deep learning* tersebut mampu melakukan klasifikasi aksara Bali secara end-to-end, serta menjadi baseline untuk dibandingkan dengan performa model gabungan. Metode gabungan yang digunakan adalah metode *Convolutional Neural Network (CNN)* sebagai pengekstrak ciri dan pengolah fitur dan digabungkan dengan *Support Vector Machine (SVM)* sebagai alat untuk klasifikasi dari data yang digunakan (Nugroho, 2021).

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang di atas, maka dapat diidentifikasi beberapa masalah sebagai berikut.

- a. Masyarakat Bali jarang membaca lontar karena dianggap sakral, sehingga sulit untuk memperoleh informasi dari lontar.
- b. Banyaknya manuskrip lontar yang belum didigitalisasi secara optimal, menyebabkan keterbatasan dalam pelestarian aksara Bali.

1.3 Pembatasan Masalah

Berdasarkan identifikasi masalah yang telah dipaparkan di atas, maka batasan masalah dalam penelitian ini adalah sebagai berikut.

- a. Metode yang digunakan adalah penggabungan *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)*.
- b. Sumber data yang digunakan dalam penelitian ini berasal dari tantangan 3: Isolated Character Recognition of Balinese Script in Palm Leaf Manuscript Images salah satu bagian AMADI_LontarSet. Set data bisa diunduh di ICFHR 2016 Challenge 3.
- c. Metrik evaluasi utama terbatas pada akurasi, presisi, recall, F1-score.
- d. Model ini akan ditujukan untuk mendukung para peneliti filologi, serta pengembang aplikasi pelestarian budaya dalam melakukan digitalisasi, klasifikasi, dan pencarian informasi berbasis aksara Bali.

1.4 Rumusan Masalah

Berdasarkan uraian yang telah dipaparkan pada latar belakang, maka rumusan masalah yang diajukan dalam penelitian ini adalah :

- a. Bagaimana arsitektur dari metode gabungan *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)* dalam mengklasifikasikan Aksara Bali dari citra Manuskrip Lontar?
- b. Bagaimana hasil akurasi model klasifikasi Aksara Bali dari citra Manuskrip Lontar yang dikembangkan dengan metode *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)*?

1.5 Tujuan Penelitian

Dilihat dari rumusan masalah di atas, adapun tujuan penelitian ini dilakukan adalah sebagai berikut.

- a. Untuk membuat model dari gabungan metode *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)* dalam klasifikasi Aksara Bali dari citra Manuskrip Lontar.
- b. Untuk menganalisis hasil akurasi klasifikasi yang menggunakan gabungan metode *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)*. serta memberikan rekomendasi untuk meningkatkan akurasi klasifikasi.

1.6 Manfaat Penelitian

a. Manfaat Praktis

1) Pelestarian Warisan Budaya

Model ini diharapkan dapat membantu proses pengenalan dan klasifikasi aksara lontar secara otomatis, mempercepat digitalisasi, serta meningkatkan pemahaman isi lontar oleh generasi muda yang kurang akrab dengan aksara Bali.

2) Implementasi Teknologi Machine Learning

Hasil penelitian ini dapat digunakan untuk mengembangkan aplikasi praktis yang membantu dalam pengenalan aksara Bali secara otomatis, memudahkan proses digitalisasi manuskrip lontar.

b. Manfaat Teoritis

1) Kontribusi pada Ilmu Pengetahuan

Penelitian ini berkontribusi pada pengembangan teknologi machine learning untuk pengenalan karakter dalam manuskrip kuno, khususnya aksara Bali, dengan menggabungkan metode .

2) Pengembangan Model Klasifikasi

Penelitian ini menghasilkan kombinasi baru antara metode *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)* dalam klasifikasi aksara, yang dapat dikembangkan lebih lanjut oleh peneliti lain untuk pengenalan aksara atau karakter dari berbagai bahasa kuno lainnya.

