BAB I
PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terakhir, transformasi teknologi informasi global yang dipicu oleh Revolusi Industri 4.0 telah mendorong kemajuan signifikan pada bidang kecerdasan buatan (*Artificial Intelligence*), khususnya dalam pengolahan citra dan video. Didukung oleh ketersediaan *Big Data* dan akselerasi komputasi GPU, teknologi visi komputer kini berevolusi menjadi sistem generatif yang mampu memanipulasi realitas visual dengan presisi tinggi. Namun, kecanggihan ini membawa dampak negatif dengan munculnya teknologi *deepfake*, sebuah teknik manipulasi video yang memanfaatkan *Deep Learning* untuk menggantikan wajah seseorang secara realistis. Fenomena ini telah meruntuhkan paradigma verifikasi visual konvensional dan menimbulkan ancaman serius terhadap integritas informasi digital [1]. Teknologi ini telah memicu krisis integritas terhadap informasi digital melalui kemampuannya dalam menciptakan manipulasi video wajah yang sangat realistis hingga sulit dibedakan dari rekaman asli. Fenomena ini membawa berbagai permasalahan serius, mulai dari penyebaran disinformasi massal yang mengancam stabilitas sosial dan politik, risiko penipuan identitas yang merugikan secara finansial, hingga pencemaran nama baik yang merusak reputasi individu secara masif. Dengan meruntuhnya paradigma "*seeing is believing*" (melihat adalah mempercayai), masyarakat kini menghadapi tantangan besar dalam memverifikasi kebenaran konten visual di media digital. Penyalahgunaan algoritma generatif ini tidak hanya mengancam privasi dan keamanan individu, tetapi juga mengikis kepercayaan publik terhadap bukti-bukti visual yang selama ini dianggap sebagai landasan kebenaran autentik.

Secara teknis, teknologi ini menyalahgunakan arsitektur *Generative Adversarial Networks* (GANs) dan *Autoencoders* untuk menukar wajah seseorang dengan wajah orang lain. Dalam mekanisme GANs, *Generator* dan *Discriminator* saling berkompetisi dalam proses pelatihan adversarial untuk menghasilkan citra tiruan yang sangat halus hingga sulit dibedakan dari aslinya. Implikasi

penyalahgunaan teknologi ini sangat luas, mulai dari penyebaran disinformasi politik, penipuan identitas, hingga pencemaran nama baik [2] , [3]. Sebagai contohnya adalah kasus video *deepfake* Presiden Prabowo Subianto dan Gibran Rakabuming Raka yang disalahgunakan penipu untuk menjanjikan bantuan bantuan sosial palsu, serta manipulasi video Menteri Keuangan Sri Mulyani yang memberikan pernyataan hoaks mengenai kebijakan anggaran negara. Di kancah internasional, manipulasi video Presiden Ukraina Volodymyr Zelenskyy yang seolah menyerukan penyerahan diri menjadi bukti nyata bagaimana *deepfake* dapat mengancam stabilitas geopolitik. Tidak hanya menyasar tokoh politik, teknologi ini juga secara masif digunakan untuk melakukan penipuan komersial dan penipuan identitas dengan mengeksploitasi popularitas tokoh terkenal. Modus penipuan berupa *giveaway* palsu atau promosi investasi bodong yang mencatut wajah selebriti seperti Raffi Ahmad dan Baim Wong, serta penggunaan citra publik figur internasional seperti Elon Musk dalam skema penipuan mata uang kripto, telah menyebabkan kerugian finansial yang signifikan bagi banyak korban. Fenomena ini mempertegas bahwa *deepfake* telah menjadi instrumen kejahatan siber yang sangat berbahaya karena mampu mengeksploitasi kepercayaan masyarakat melalui manipulasi visual yang sangat meyakinkan.

Tantangan pendeteksian menjadi semakin kompleks karena perkembangan terbaru menampilkan video *deepfake* yang semakin sulit dibedakan oleh mata manusia maupun metode deteksi konvensional. Metode generasi awal yang mengandalkan analisis artefak visual permukaan, seperti ketidakselarasan resolusi atau efek *blur*, mulai kehilangan relevansinya [4]. Hal ini disebabkan oleh kemampuan model modern dalam melakukan *refinement* otomatis, serta pengaruh ekosistem distribusi konten digital [5]. Seiring meningkatnya kualitas *deepfake*, metode deteksi berbasis artefak visual menjadi kurang efektif karena video yang beredar di media sosial umumnya telah mengalami kompresi *lossy* atau *post-processing* halus. Proses kompresi ini menghapus detail frekuensi tinggi yang biasanya menjadi jejak digital manipulasi, menyebabkan kegagalan deteksi pada video *in-the-wild*.

Tingkat realisme model generatif saat ini, yang mampu mensimulasikan pencahayaan dinamis dan tekstur kulit mendetail, menuntut adanya pergeseran pendekatan deteksi. Integritas visual tidak lagi dapat diandalkan sepenuhnya, sehingga metode verifikasi harus beralih ke aspek yang lebih fundamental: aspek biologis. Salah satu pendekatan yang dinilai menjanjikan adalah deteksi berbasis anomali fisiologis atau *passive liveness detection*. Hipotesis utamanya adalah bahwa meskipun AI mampu meniru "penampilan" fisik, ia kesulitan mereplikasi fungsi fisiologis otonom manusia secara akurat.

Pendekatan ini memanfaatkan sinyal biologis alami, salah satunya adalah sinyal *Remote Photoplethysmography* (rPPG). rPPG bekerja dengan mendeteksi perubahan warna kulit yang sangat halus akibat aliran darah yang dipompa oleh jantung. Hemoglobin dalam darah menyerap cahaya hijau lebih kuat dibandingkan spektrum lain, menciptakan fluktuasi warna ritmis yang sinkron dengan detak jantung. Fitur ini secara alami ada pada video wajah asli, namun sering kali absen, datar, atau tidak konsisten pada video hasil sintesis *deepfake* karena generator AI umumnya merekonstruksi wajah *frame-by-frame* tanpa memperhitungkan koherensi fisiologis antar-frame.

Meskipun menawarkan landasan ilmiah yang kuat, penggunaan modalitas rPPG tunggal memiliki keterbatasan. Sinyal rPPG sangat sensitif terhadap gangguan eksternal seperti perubahan pencahayaan ekstrem dan pergerakan kepala yang signifikan, yang dapat menurunkan rasio *Signal-to-Noise* (SNR). Lebih krusial lagi, studi keamanan terbaru menunjukkan adanya ancaman serangan manipulasi sinyal fisiologis, seperti *PulseEdit*, yang mencoba mengelabui detektor rPPG dengan menyuntikkan sinyal detak jantung palsu ke dalam video sintetis. Ancaman ini menegaskan bahwa sistem keamanan yang bersifat unimodal (satu fitur) tidak lagi memadai.

Untuk memitigasi risiko tersebut, diperlukan strategi pertahanan berlapis dengan mengintegrasikan fitur fisiologis internal (rPPG) dengan fitur perilaku fisik eksternal, yaitu pola kedipan mata. Pola kedipan mata pada video *deepfake* sering menunjukkan ketidakwajaran temporal, seperti frekuensi kedipan yang terlalu

rendah (*zombie-like gaze*) atau desinkronisasi antar-mata. Sedangkan sinyal rPPG merepresentasikan perubahan mikrovaskuler internal. Penggabungan kedua fitur ini diyakini dapat menutupi kelemahan masing-masing.

Berdasarkan kondisi tersebut, penelitian ini difokuskan pada deteksi *deepfake* pada video wajah dengan memanfaatkan analisis anomali fisiologis berbasis sinyal rPPG dan pola kedipan mata. Sistem yang dirancang menerapkan fusi multimodal menggunakan pendekatan *Deep Learning* (seperti LSTM) untuk memodelkan karakteristik temporal dari kedua fitur tersebut [6]. Melalui pendekatan ini, diharapkan sistem mampu meningkatkan akurasi klasifikasi antara video wajah asli dan video *deepfake* secara lebih presisi, serta memiliki ketahanan (*robustness*) yang lebih baik terhadap variasi kualitas video maupun upaya manipulasi sinyal di dunia nyata.

1.2 Rumusan Masalah

Berdasar latar belakang masalah seperti di 1.1, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang bangun sistem deteksi deepfake yang mengintegrasikan ekstraksi fitur sinyal fisiologis *remote Photoplethysmography* (rPPG) dan fitur geometris *Eye Aspect Ratio* (EAR) menggunakan *library* MediaPipe?
2. Seberapa tinggi tingkat akurasi model yang diusulkan dalam mendeteksi video deepfake dibandingkan dengan penggunaan fitur tunggal?
3. Bagaimana efektif pendekatan anomali fisiologis dalam membedakan video wajah asli dan video *deepfake*?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan terfokus, maka batasan masalah yang diterapkan adalah sebagai berikut:

1. Objek penelitian berupa video wajah tunggal dengan klasifikasi *real* dan *deepfake*.

2. Fitur yang dianalisis dibatasi pada sinyal fisiologis rPPG (menggunakan *Green Channel*) dan pola kedipan mata (menggunakan *Eye Aspect Ratio*/EAR) yang diekstraksi menggunakan *library* MediaPipe.
3. Metode klasifikasi menggunakan pendekatan *deep learning* berbasis data sekuens.
4. Penelitian tidak membahas aspek audio maupun *real-time system*.
5. Dataset yang digunakan merupakan dataset sekunder yang diperoleh dari sumber publik.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian deteksi video deepfake menggunkan pendekatan anomali fisiologis adalah sebagai berikut:

1. Menganalisis perbedaan karakteristik sinyal rPPG dan pola kedipan mata antara video wajah asli dan video *deepfake*.
2. Merancang model *deep learning* untuk mendeteksi *deepfake* berdasarkan fitur fisiologis rPPG (*remote Photoplethysmography*) dan pola kedipan mata (EAR) menggunakan *library* MediaPipe.
3. Mengukur efektivitas pendekatan anomali fisiologis dalam mendeteksi dan klasifikasi video *deepfake*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian deteksi video deepfake menggunkan pendekatan anomali fisiologis adalah sebagai berikut:

1. Manfaat secara teoritis adalah untuk memberikan kontribusi terhadap pengembangan kajian deteksi *deepfake* berbasis sinyal fisiologis dan *liveness detection*.
2. Manfaat secara praktis adalah verifikasi keaslian video digital untuk membantu masyarakat atau institusi dalam meminimalisir penyebaran

hoaks dan disinformasi yang di buat dengan *deepfake* yang fokus berbasis pada sinyal fisiologis dan *liveness detection*.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran alur pembahasan pada setiap bab.

Bab I pendahuluan, berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

Bab II Tinjauan Pustaka memuat studi literatur dan dasar teori yang berkaitan dengan *deepfake*, rPPG, pola kedipan mata, serta *deep learning*.

Bab III Metode Penelitian menjelaskan objek penelitian, alur penelitian, serta alat dan bahan yang digunakan.

Bab IV Hasil dan Pembahasan menyajikan hasil eksperimen dan analisis kinerja metode yang diusulkan.

Bab V Penutup berisi kesimpulan dan saran untuk pengembangan penelitian selanjutnya.