

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam bisbol, melempar bola (*pitching*) adalah salah satu gerakan yang paling menuntut fisik dan rumit secara teknis di semua olahraga profesional [1]. Performa seorang pelempar (*pitcher*) menentukan hasil pertandingan dan musim kompetisi, namun juga memberikan tekanan berulang yang berat pada lengan, terutama ligamen kolateral ulnaris (UCL) [2]. Cedera yang diakibatkannya telah menjadi kasus yang meningkat dalam bisbol saat ini, sedemikian parahannya sehingga para ahli kini menggambarkannya sebagai “epidemi” operasi rekonstruksi [3].

Rekonstruksi UCL, yang umumnya dikenal sebagai *Tommy John Surgery* (TJS), semakin lazim di kalangan pelempar *Major League Baseball* (MLB), dengan lebih dari 35–37% pelempar aktif telah menjalani prosedur ini setidaknya sekali [4], [5]. Selain implikasi medis, beban ekonominya juga signifikan, dengan absennya pemain akibat TJS yang merugikan tim MLB ratusan juta dolar setiap tahunnya akibat hilangnya produktivitas dan investasi rehabilitasi [6].

Jumlah prosedur TJS telah meningkat secara eksponensial dalam dekade terakhir. Pada tahun 2024 saja, MLB mencatat lebih banyak operasi *Tommy John* daripada di seluruh dekade 1990-an [7], menggarisbawahi regangan pada pelempar yang belum pernah terjadi sebelumnya di era modern. Ahli bedah menghubungkan hal ini dengan budaya bisbol yang semakin terobsesi dengan kecepatan lemparan (*velocity*) dan kecepatan putaran (*spin rate*) sebagai tuntutan mekanis yang mengintensifkan regangan pada UCL. Dr. Timothy Kremchek, mantan direktur medis *Cincinnati Reds*, mencatat peningkatan frekuensi operasi *Tommy John* yang drastis, di mana jumlah prosedur yang dahulu dilakukan dalam satu tahun kini sejumlah yang pernah dia tangani dalam kurun waktu satu minggu saja [3]. Sejak rekonstruksi UCL pertama dilakukan pada 25 September 1974, prosedur inovatif ini telah menjadi salah satu operasi paling berpengaruh dalam kedokteran olahraga [8].

Secara historis, upaya untuk mencegah cedera terkait *pitching* berpusat pada dua pendekatan umum. Pendekatan pertama adalah pemantauan beban kerja (*workload monitoring*), yang melacak babak (*inning*) yang dilempar, jumlah total lemparan, atau hari istirahat. Meskipun intuitif, pengukuran ini seringkali

gagal menangkap stres biomekanik yang mendasari regangan ligamen [9], [10]. Pendekatan kedua melibatkan analisis biomekanik di laboratorium, biasanya menggunakan sistem tangkap gerak (*motion capture*) untuk mengukur torsi sendi, kinematika lengan, dan beban otot [1], [11]. Meskipun studi biomekanik telah memperdalam pemahaman tentang mekanika lemparan, keterbatasan eksperimental dan ukuran sampel yang terbatas membatasi kemampuannya untuk digeneralisasi ke kondisi dunia nyata dan tingkat permainan MLB [12].

Perubahan mendasar terjadi dengan diperkenalkannya *Statcast*, sistem pelacakan optik canggih MLB yang diperkenalkan pada 2015 [13]. *Statcast* menyediakan pengukuran kinematik berfrekuensi tinggi untuk setiap lemparan dalam pertandingan MLB—meliputi kecepatan pelepasan, laju putaran (*spin rate*), ekstensi, posisi pelepasan, dan metrik pergerakan bola [14]. Metrik-metrik ini merepresentasikan hasil luar dari mekanika tubuh pelempar (*output kinematics*) dan berfungsi sebagai indikator tidak langsung terhadap perubahan biomekanik yang mendasarinya. Teknologi *Statcast* memungkinkan analisis kinematik berskala besar dalam kondisi pertandingan, menjembatani kesenjangan antara presisi laboratorium dan penerapan di lapangan [15]. Dengan demikian, peneliti kini dapat mempelajari bagaimana variasi dalam konsistensi mekanis dari ribuan lemparan dapat menandakan risiko cedera yang muncul [16]. Ketersediaan data ini juga telah memicu generasi baru penelitian berbasis data dan pembelajaran mesin mengenai prediksi cedera.

Oeding et al. (2024) menganalisis metrik pelacakan lemparan menggunakan *Extreme Gradient Boosting (XGBoost)* dengan interpretabilitas SHAP untuk memprediksi cedera bahu dan siku di masa mendatang pada pelempar MLB. Studi mereka menunjukkan bahwa fluktuasi dalam kecepatan, laju putaran (*spin rate*), dan pola penggunaan lemparan dapat mengindikasikan risiko cedera dini, menegaskan bahwa kerangka kerja pembelajaran mesin yang dapat dijelaskan (*explainable machine learning*) mampu mengekstrak wawasan kinematik dari data pelacakan skala besar. Namun, analisis mereka bergantung pada rata-rata tingkat musim dan tidak memeriksa fluktuasi jangka pendek dalam mekanika sebelum cedera. Keterbatasan ini menyoroti perlunya model *time-resolved* yang mampu memodelkan perubahan mekanis pelempar yang dilacak dari waktu ke waktu [17].

Penelitian selanjutnya seperti Butani et al. (2025) menerapkan model *tabular interpretable*, termasuk *Deep Neural Decision Forest (DNDF)*, untuk mengklasifikasikan risiko cedera UCL berdasarkan beban kerja agregat dan fitur performa. Meskipun model mereka efisien dan mudah ditafsirkan, penelitian

tersebut tidak memanfaatkan data rinci maupun *pitch-by-pitch*, sehingga kurang mendalam secara temporal dan kinematik [18].

Kang et al. (2025) kemudian memperkenalkan kerangka kerja pembelajaran mendalam yang memanfaatkan data *Statcast* per lemparan (*pitch-by-pitch data*) untuk memprediksi risiko TJS. Kang menerapkan arsitektur sekuensial, termasuk *Long Short-Term Memory* (LSTM), *Convolutional Neural Network* (CNN), dan *Vision Transformer* (ViT), dalam rentang waktu 100 hari sebelum cedera. Model ViT mereka mencapai akurasi prediktif tinggi serta hasil terbaik di antara model yang diuji, dan menangkap degradasi *spasial-temporal* bertahap dari mekanika lemparan yang sering kali mendahului kerusakan UCL [19].

Meskipun terdapat kemajuan-kemajuan ini, studi yang ada masih menyoroti kesenjangan yang jelas. Kerangka kerja Oeding dan Butani menekankan interpretabilitas tetapi tidak mampu memodelkan evolusi mekanis dari waktu ke waktu, sementara arsitektur ViT Kang menawarkan prediksi temporal kuat namun terbatas pada data MLB. Kang mencatat bahwa kemajuan di masa mendatang memerlukan perluasan kumpulan data baik dari *Minor League Baseball* (MiLB) maupun liga internasional, dengan harapan dapat meningkatkan keandalan dan generalisasi model sehingga memastikan penerapan yang lebih luas di berbagai tingkat persaingan [19].

Berdasarkan wawasan dan keterbatasan yang diidentifikasi dalam studi sebelumnya, penelitian ini mengembangkan pendekatan prediksi cedera berbasis data yang lebih komprehensif. Penelitian ini memperluas cakupan analisis dengan mengintegrasikan data *Statcast* dari *Minor League Baseball* (MiLB)—khususnya pada tingkat MiLB tertinggi yaitu *Triple-A* (AAA)—yang dipilih karena merupakan satu-satunya level liga minor dengan cakupan *Statcast* penuh di semua tim dan stadion sejak 2023 [20]. Penyertaan ini memungkinkan evaluasi generalisasi model di berbagai level kompetisi dengan tetap menjaga konsistensi kualitas dan presisi pengukuran. Lebih lanjut, penelitian ini menerapkan teknik agregasi temporal dalam berbagai rentang waktu melalui analisis sensitivitas jendela-bergulir (*rolling-window sensitivity analysis*) dengan durasi 20, 40, 60, 80, hingga 100 hari. Metode ini bertujuan untuk menangkap tren statistik dan variabilitas mekanis pelempar pada berbagai tingkat resolusi temporal, untuk mengidentifikasi jendela waktu yang paling stabil dan prediktif dalam mendeteksi sinyal cedera dibandingkan analisis lemparan tunggal.

Dengan mengadopsi arsitektur ViT sebagai kerangka kerja utama dan menerapkannya pada data agregat lintas level (MLB dan AAA) ini, penelitian

ini bertujuan untuk membangun pendekatan yang lebih *robust*, kontekstual, dan dapat meningkatkan generalisasi model sehingga memastikan penerapan yang lebih luas di berbagai tingkat persaingan profesional, serta memberikan interpretabilitas model yang jelas [21].

1.2 Rumusan Masalah

Eskalasi angka cedera *Ulnar Collateral Ligament* (UCL) di era bisbol modern mengindikasikan adanya tekanan mekanis sistemik yang melampaui batas adaptasi fisik pelempar profesional. Meskipun sistem *Statcast* menyediakan data kinematik resolusi tinggi, terdapat celah besar dalam menginterpretasikan data tersebut untuk menangkap sinyal degradasi fisik yang bersifat akumulatif dan volatil, terutama ketika melibatkan pemain dalam fase pengembangan di level liga yang berbeda. Oleh karena itu, diperlukan sebuah pendekatan *deep learning* yang mampu mengekstraksi pola spasial-temporal dari data lintas liga untuk mengidentifikasi indikator terbesar sebelum terjadinya cedera UCL. Berdasarkan uraian tersebut, masalah dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana pengembangan model *Vision Transformer* dalam mengintegrasikan data lintas level (MLB dan AAA) untuk mengidentifikasi fitur kinematik yang menjadi indikator terkuat risiko cedera UCL?
2. Bagaimana efektivitas berbagai rentang jendela waktu (*time series*) dalam memodelkan perubahan biomekanik untuk menghasilkan prediksi risiko cedera yang paling akurat berdasarkan metrik evaluasi?

1.3 Batasan Permasalahan

1. Penelitian ini berfokus pada *pitcher* profesional yang datanya tercatat dalam sistem *Statcast Major League Baseball* (MLB) dan *Triple-A* (AAA).
2. Data bersumber dari *Statcast* (melalui *pybaseball*) dan mencakup data musim reguler 2017 hingga 2025. Tanggal mulai 2017 dipilih untuk memastikan konsistensi dan kualitas data *Statcast*, di mana tahun 2016 dan sebelumnya masih menggunakan *Pitch F/X* yang belum seakurat *Statcast*. Data terbatas pada liga bisbol Amerika karena liga internasional tidak menyediakan data rinci dan mendalam secara publik.

3. *Pitcher* dikelompokkan menjadi dua kategori berdasarkan status cedera: (1) *pitcher* yang menjalani operasi *Tommy John Surgery* (TJS), dan (2) *pitcher* kontrol yang tidak mengalami cedera.

Kriteria untuk Kelompok Cedera (TJS) adalah:

- (a) Berasal dari daftar publik catatan kasus operasi TJS pada pemain bisbol profesional (Roegel).
- (b) Level *pitcher* yang tercatat dalam daftar tersebut adalah 'MLB' atau 'AAA'.
- (c) Tanggal operasi TJS harus terjadi pada atau setelah 1 Januari 2017.

Kriteria untuk Kelompok Kontrol (Sehat) adalah:

- (a) *Pitcher* tidak boleh terdaftar dalam daftar TJS.
 - (b) Untuk level MLB, *pitcher* harus memenuhi kriteria partisipasi signifikan dalam 3 dari 4 musim berturut-turut sesuai logika Kang (2025).
 - (c) Untuk level AAA, diterapkan kriteria *workhorse* dengan ambang batas beban kerja minimal 700 lemparan per musim selama setidaknya 2 musim dalam rentang 2023–2025.
 - (d) Penentuan Titik Acuan (*Anchoring*): Untuk kelompok kontrol, tanggal pertandingan terakhir ditetapkan sebagai "Day 0" (*anchor*) untuk menghitung mundur jendela waktu performa, sehingga menciptakan konsistensi temporal dengan data pra-cedera pada kelompok TJS.
4. *Dataset* diubah dari data *pitch-by-pitch* mentah menjadi data agregat jendela-bergulir (*rolling-window aggregate*) yang diuji secara komparatif pada rentang 20, 40, 60, 80, dan 100 hari untuk menentukan resolusi temporal optimal dalam prediksi risiko cedera.
 5. Fitur *Statcast* dengan nilai hilang (*missing values* atau NaN) lebih dari 90
 6. Faktor non-mekanik seperti kondisi psikologis, beban latihan di luar pertandingan, atau faktor lingkungan tidak termasuk dalam cakupan penelitian.

1.4 Tujuan Penelitian

1. Mengembangkan model *Vision Transformer* yang mengintegrasikan data lintas level (*MLB* dan *AAA*) serta mengidentifikasi fitur kinematik *Statcast* yang menjadi indikator terkuat risiko cedera UCL melalui teknik analisis interpretabilitas bertingkat (*Attention Map* dan korelasi statistik *Pearson*).
2. Mengevaluasi efektivitas berbagai rentang jendela waktu antara 20 hingga 100 hari dalam memodelkan perubahan biomekanik untuk menghasilkan prediksi risiko cedera UCL yang paling optimal berdasarkan metrik evaluasi klasifikasi.

1.5 Manfaat Penelitian

1. Memberikan kerangka kerja berbasis data untuk deteksi dini risiko cedera UCL pada pelempar (*pitcher*) bisbol profesional, yang dapat dimanfaatkan oleh pelatih, analis performa, maupun tim medis.
2. Menjadi kontribusi ilmiah dalam penerapan model *Vision Transformer* (ViT) dan pendekatan *multimodal* pada analisis biomekanik olahraga.
3. Mendorong pemanfaatan data *Statcast* level *Minor League* (*AAA*) untuk memperluas cakupan riset biomekanik berbasis *machine learning* di bidang bisbol.
4. Menjadi dasar bagi penelitian lanjutan dalam pengembangan sistem pencegahan cedera dan manajemen beban kerja (*workload management*) atlet berbasis data prediktif.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN Menjelaskan latar belakang masalah, rumusan masalah, batasan penelitian, tujuan, manfaat, serta sistematika penulisan laporan.
- Bab 2 LANDASAN TEORI Membahas teori-teori yang mendasari penelitian, termasuk konsep biomekanika *pitching*, cedera UCL, sistem *Statcast*, teknik

rekayasa fitur (*feature engineering*) untuk data deret waktu, penanganan *dataset* tidak seimbang, serta model pembelajaran mesin yang digunakan yaitu *Vision Transformer* (ViT).

- Bab 3 METODOLOGI PENELITIAN Menguraikan rancangan penelitian, prosedur pembangunan *dataset* MLB dan AAA, serta tahapan pemodelan prediktif dan metrik evaluasi yang digunakan.
- Bab 4 HASIL DAN DISKUSI Menyajikan hasil eksperimen model, analisis perbandingan performa antara model *baseline* dan ViT, serta pembahasan mendalam mengenai fitur kinematik melalui pemetaan atensi internal (*attention map*), arah fisik perubahan metrik dengan korelasi *Pearson*, serta analisis komparatif manajemen risiko antara level liga MLB dan AAA.
- Bab 5 KESIMPULAN DAN SARAN Menyimpulkan temuan utama penelitian serta memberikan saran strategis untuk penelitian lanjutan.

