

## MODEL PREDIKSI TABRAKAN BERBASIS MACHINE LEARNING DALAM SISTEM INTERNET OF VEHICLES (IoV)

Raden Teddy Iswahyudi <sup>1</sup>, Nizirwan Anwar <sup>2</sup>

<sup>1,2</sup> Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Esa Unggul

raden.teddy@esaunggul.ac.id  
nizirwan.anwar@esaunggul.ac.id

### ABSTRAK

Perkembangan Internet of Vehicles (IoV) menghadirkan peluang signifikan dalam peningkatan keselamatan berkendara melalui sistem prediksi tabrakan berbasis data. Penelitian ini bertujuan untuk mengembangkan model prediktif yang akurat dan efisien dengan menggunakan pendekatan machine learning, khususnya algoritma klasifikasi seperti Support Vector Machine (SVM), Random Forest, dan Multi Layer Perceptron (MLP). Dataset yang digunakan berasal dari Kaggle dan mencakup 207.000 entri dengan fitur-fitur seperti kecepatan kendaraan, jarak antar kendaraan, status pengemudi, kondisi lingkungan, dan atribut jaringan komunikasi. Tahapan penelitian meliputi eksplorasi data, preprocessing, seleksi fitur, pelatihan model, dan evaluasi performa menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan performa model yang sangat tinggi, yaitu akurasi 99,99%, presisi 99,99%, recall 100%, dan F1-score 99,99%. Evaluasi confusion matrix menunjukkan bahwa hampir semua kasus diklasifikasikan dengan benar, dan nilai AUC ROC curve sebesar 0,9999999 menandakan kemampuan diskriminasi yang hampir sempurna. Fitur paling berpengaruh dalam prediksi tabrakan adalah status pengemudi, jumlah jalur, dan kapabilitas pengereman. Temuan ini menunjukkan bahwa model berbasis machine learning mampu mendeteksi potensi tabrakan secara tepat dan real-time, serta dapat diintegrasikan ke dalam sistem peringatan dini pada kendaraan cerdas, menjadikannya solusi potensial untuk mengurangi risiko kecelakaan dalam ekosistem IoV yang dinamis dan kompleks.

**Kata kunci:** Internet of Vehicles, Prediksi Tabrakan, Machine Learning, Feature Selection, SVM

### 1. PENDAHULUAN

Perkembangan teknologi komunikasi dalam sektor transportasi telah mendorong kemunculan konsep **Internet of Vehicles (IoV)**, yaitu ekosistem kendaraan yang saling terkoneksi melalui jaringan data untuk bertukar informasi secara real-time. IoV tidak hanya meningkatkan kenyamanan dan keselamatan dalam berkendara, tetapi juga menjadi solusi potensial dalam menekan angka kecelakaan lalu lintas melalui sistem prediksi dan peringatan dini berbasis data (Chang et al., 2021; Qian, 2022). Dalam konteks ini, **prediksi tabrakan** menjadi salah satu komponen vital yang membutuhkan pemodelan cerdas berbasis data dinamis, seperti kecepatan kendaraan, jarak antar kendaraan, status pengemudi, kondisi lingkungan, dan pola lalu lintas (Almutairi et al., 2023a; Yang et al., 2022). Namun demikian, implementasi sistem prediksi tabrakan masih diemikan dan menghadapi sejumlah tantangan dan peluang,

seperti **pemilihan fitur yang relevan, ketersediaan data yang berkualitas, dan pemilihan algoritma yang tepat** untuk mengolah data kompleks dan heterogen secara efisien (Manchala & Vijaya Kishore, 2024). Selain itu, kemampuan sistem dalam memberikan respons secara real-time dalam kondisi lalu lintas yang dinamis menjadi faktor penentu keberhasilan sistem prediksi tabrakan yang efektif. Identifikasi faktor-faktor yang paling berkontribusi terhadap terjadinya tabrakan menjadi penting untuk meningkatkan akurasi model dan efektivitas sistem penghindaran tabrakan (Jeong et al., 2020). Urgensi dari penelitian ini pada meningkatnya ketergantungan kendaraan terhadap sistem berbasis data (datasets) untuk mendukung keamanan. Dalam perubahan global dunia yang semakin terotomatisasi, sistem prediksi berbasis IoV bukan lagi pilihan, melainkan kebutuhan mendesak. Oleh karena itu, dibutuhkan pemodelan yang mampu mengakomodasi kondisi nyata di lapangan

dengan mempertimbangkan parameter teknis, perilaku, dan lingkungan secara simultan. Rumusan masalah yang dikaji dalam studi ini adalah: “*Bagaimana memodelkan sistem prediksi tabrakan dalam lingkungan IoV dengan menggunakan data dinamis secara akurat dan efisien?*”. Tujuan utama dari penelitian ini adalah mengembangkan model prediksi tabrakan berbasis data yang mempertimbangkan berbagai parameter kendaraan dan lingkungan dengan pendekatan machine learning untuk meningkatkan keselamatan berkendara. Adapun manfaat dari penelitian ini mencakup dua hal utama: (1) **Secara teoritis**, memberikan kontribusi pada pengembangan model prediktif berbasis IoV yang lebih akurat dan adaptif terhadap dinamika lalu lintas; (2) **Secara praktis**, menyediakan dasar implementasi sistem peringatan dini tabrakan yang dapat diintegrasikan dalam kendaraan cerdas dan sistem lalu lintas masa depan.

## 2. METODOLOGI

### 2.1 Teori Pendukung

Penelitian ini berakar pada pendekatan *machine learning* dalam konteks prediksi tabrakan pada sistem **Internet of Vehicles (IoV)**. Teori utama yang digunakan adalah teori *supervised learning*, khususnya *classification algorithms*, yang digunakan untuk membedakan kondisi lalu lintas "berisiko tabrakan" dan "tidak berisiko". Algoritma yang umum digunakan dalam konteks ini termasuk **Random Forest**, **Support Vector Machine (SVM)**, dan **Neural Network**, karena kemampuannya dalam menangani data non-linear, multivariat, dan bersifat real-time lintas (Almutairi et al., 2023a; Yang et al., 2022). Pemodelan prediktif ini bertumpu pada prinsip *feature engineering*, yaitu proses memilih dan mentransformasikan fitur yang paling relevan dari data kendaraan dan jaringan komunikasi. Proses ini penting untuk meningkatkan akurasi model serta mengurangi kompleksitas komputasi (Qian,

2022). Selain itu, prinsip *data-driven decision making* digunakan untuk memastikan bahwa seluruh hasil prediksi berbasis pada representasi data aktual yang dikumpulkan dari simulasi atau data historis.

### 2.2 Desain Penelitian

Penelitian ini menggunakan pendekatan **kuantitatif eksperimental**, dengan metode simulasi dan analisis data sekunder dari dataset yang telah tersedia. Desain ini mencakup proses eksplorasi data, preprocessing, pembangunan model prediksi, evaluasi model, dan interpretasi hasil. Penelitian dibagi ke dalam beberapa tahap secara makro (*main*) sebagai berikut (gambar 1):

#### 1. Pengumpulan dan Eksplorasi Data

Dataset (**207.000 rows**) yang digunakan mencakup data lalu lintas jaringan kendaraan (Gombe, 2023) yang terdiri atas fitur-fitur seperti kecepatan, jarak antar kendaraan, status pengemudi, kondisi lingkungan, dan riwayat koneksi jaringan (protocol, port, IP, bytes, dll). Data ini dieksplorasi untuk memahami distribusi, outlier, serta relevansi fitur terhadap label target (*collision/benign*).

#### 2. Preprocessing Data

Meliputi pembersihan data (*handling missing values*), normalisasi/ standarisasi numerik, encoding fitur kategorikal, dan pembagian data ke dalam set pelatihan (*training*) dan pengujian (*testing*). Proses seleksi fitur juga dilakukan berdasarkan korelasi dan *feature importance*.

#### 3. Pemodelan Prediktif

Beberapa model klasifikasi diuji, seperti SVM, Random Forest, dan *Multi Layer Perceptron* (MLP). Model dilatih menggunakan data pelatihan dan divalidasi melalui metode *k-fold cross-validation* untuk menghindari *overfitting*.

#### 4. Evaluasi Model

Performa model diukur dengan menggunakan metrik seperti **akurasi**, **presisi**, **recall**, dan **F1-score**, dengan perhatian khusus pada *recall* sebagai indikator kemampuan model

mendeteksi potensi tabrakan secara optimal (Manchala & Vijaya Kishore, 2024).

### 5. Interpretasi dan Visualisasi

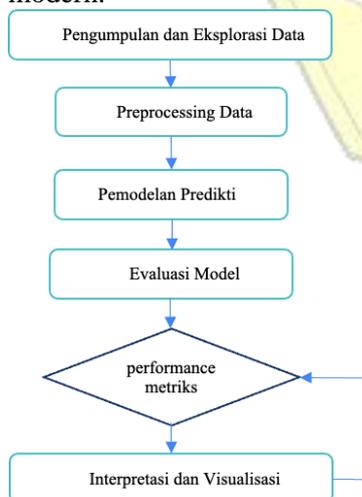
Hasil model divisualisasikan dalam bentuk confusion matrix, ROC curve, dan *heatmap* korelasi antar fitur, untuk mendukung analisis lebih lanjut mengenai pola tabrakan.

## 3. LANDASAN TEORI

### 3.1 Internet of Vehicles (IoV)

*Internet of Vehicles (IoV)* merupakan evolusi dari *Vehicular Ad-Hoc Network (VANET)* yang memungkinkan kendaraan untuk berkomunikasi dengan entitas lain seperti kendaraan lain (V2V), infrastruktur (V2I), dan pejalan kaki (V2P) melalui koneksi internet dan sistem sensor ((Qian, 2022)). IoV menciptakan ekosistem kendaraan cerdas yang mampu mengambil keputusan berdasarkan data lingkungan dan informasi lalu lintas secara real-time.

Teori dasar IoV mencakup konsep *Cyber-Physical Systems (CPS)* yang menyatukan komputasi, komunikasi, dan kontrol dalam kendaraan, serta *Edge Computing* untuk mengurangi latensi dalam pengolahan data kritikal seperti prediksi tabrakan (Chang et al., 2021). Hal ini akan bersinergi dan konstruktif peran IoV dalam sistem keamanan berkendara modern.



Gambar 1. Tahapan Penelitian

### 3.2 Prediksi Tabrakan (*Collision Prediction Theory*)

Prediksi tabrakan mengacu pada teori probabilistic yang prediksi dan evaluasi kemungkinan terjadinya benturan antar kendaraan berdasarkan parameter seperti kecepatan, arah, jarak, serta reaksi pengemudi. Teori ini berakar pada *Time to Collision (TTC)*, waktu yang dibutuhkan dua kendaraan untuk bertabrakan jika mereka melanjutkan laju saat ini. Dalam sistem IoV, TTC dikombinasikan dengan *contextual decision-making models* seperti Bayesian Networks dan fuzzy logic untuk mengantisipasi tabrakan secara lebih adaptif terhadap kondisi lingkungan (Jeong et al., 2020). Relevansi teori ini sangat tinggi karena mendasari keputusan sistem peringatan dini (*collision warning system*).

### 3.3 Machine Learning untuk Prediksi Kejadian

*Machine Learning (ML)* adalah metode pendekatan komputasional yang memungkinkan sistem untuk belajar dari data tanpa di program secara eksplisit. Dalam konteks IoV, teori dan konsep *supervised learning* digunakan untuk mengembangkan model klasifikasi antara kondisi aman (*benign*) dan berisiko tabrakan (*malicious*). Model seperti **Support Vector Machine (SVM)**, **Random Forest**, dan **Long Short-Term Memory (LSTM)** telah terbukti efektif dalam menangkap pola temporal dan spasial dari lalu lintas kendaraan (Almutairi et al., 2023b; Manchala & Vijaya Kishore, 2024). Kekuatan teori ini terletak pada kemampuan untuk menangani volume data besar dan variatif, adaptif serta dinamis terhadap perubahan pola lalu lintas secara real-time.

### 3.4 Feature Engineering dan Selection

Landasan lain yang penting adalah teori *feature engineering*, yaitu proses konstruksi dan pemilihan atribut data yang relevan dan signifikan dalam mempengaruhi output model prediksi. Teori ini berperan besar dalam meningkatkan performa model dan menghindari *overfitting* akibat penggunaan fitur yang tidak informatif (Strickland et al., 2018). Dalam konteks tabrakan, fitur seperti *velocity*, *distance*, *connection state*, dan *traffic flow direction* sangat menentukan. Oleh karena itu, penggabungan domain knowledge dan statistik eksploratif diperlukan dalam proses seleksi fitur yang optimal.

### 3.5 Performance Model SVM

Dalam evaluasi performa model klasifikasi, keempat metrik (**akurasi, presisi, recall, dan F1-score**) saling berkaitan dan berperan penting dalam memberikan gambaran yang menyeluruh tentang kualitas prediksi model. Masing-masing metrik memiliki fokus evaluatif yang berbeda, namun secara kolektif membentuk dasar analisis performa yang kuat. **Akurasi** merupakan metrik paling umum dan sederhana, menggambarkan proporsi total prediksi yang benar terhadap seluruh prediksi. Meskipun sering digunakan sebagai indikator utama keberhasilan model, akurasi **bisa menyesatkan** jika digunakan secara tunggal, terutama pada dataset yang **tidak seimbang** (misalnya, kasus tabrakan yang jauh lebih sedikit daripada kondisi aman). Dalam skenario ini, model dapat mencapai akurasi tinggi hanya dengan memprediksi semua data sebagai kondisi "aman", tanpa benar-benar mendeteksi tabrakan.

Oleh karena itu, **presisi dan recall** menjadi metrik pelengkap yang lebih informatif. **Presisi** menunjukkan sejauh mana model akurat dalam prediksi positifnya dari seluruh prediksi bahwa tabrakan akan terjadi, berapa banyak yang benar-benar terjadi. Ini penting untuk menghindari **false positives**, yang dalam sistem nyata dapat menyebabkan peringatan palsu dan mengganggu pengalaman pengguna.

Di sisi lain, **recall** atau sensitivitas berfokus pada kemampuan model dalam menangkap semua kejadian positif yang sebenarnya untuk men-deteksi semua insiden tabrakan yang benar-benar terjadi. Recall tinggi berarti **false negatives** dapat diminimalkan, yang sangat penting untuk sistem keamanan yang responsif. Namun, ada kalanya terjadi **trade-off antara presisi dan recall**. Misalnya, meningkatkan recall dapat menyebabkan presisi menurun, dan sebaliknya. Untuk itu, **F1-score** sebagai metrik kompromi yang menggabungkan keduanya dalam satu angka harmonis. F1-score tinggi menandakan bahwa model tidak hanya memiliki presisi yang baik, tetapi juga tidak mengorbankan kemampuannya dalam mendeteksi kasus positif. Secara umum, **akurasi** memberi pandangan global, **presisi** fokus pada kebenaran prediksi positif, **recall** fokus pada kelengkapan deteksi, dan **F1-score** menjembatani keduanya. Keempat metrik ini, jika dianalisis secara bersamaan, akan memberikan **gambaran utuh dan seimbang** mengenai performa model, baik dalam konteks umum maupun aplikasi kritikal seperti prediksi kecelakaan dalam IoV.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Rancangan Model Prediksi

Rancangan dan tahapan dalam membangun Model Prediksi untuk kasus prediksi tabrakan dalam konteks IoV menggunakan data dinamis, sesuai dengan prinsip *machine learning pipeline* dan praktik standar riset terkini:

**Tujuan:** memprediksi kemungkinan terjadinya tabrakan berdasarkan data iov (kecepatan, jarak, status pengemudi, kondisi lingkungan, dll.).

**Tipe masalah:** *Binary classification* → Tabrakan (1) atau Tidak (0).

**Dataset:** Kaggle IoV Dataset (Gombe, 2023)  
**Selection and Training** (Pemilihan dan Pelatihan) Model; rekomendasi:

- Logistic Regression (baseline)

- Random Forest
  - Support Vector Machine (SVM)
- Training:** menggunakan data latih untuk membangun model.
- Evaluasi Performa** menggunakan metrik:
- Akurasi (keseluruhan prediksi benar)
  - Presisi (ketepatan prediksi tabrakan)
  - Recall (kemampuan mendeteksi semua tabrakan)
  - F1-score (komposisi antara presisi dan recall)
  - Confusion Matrix & ROC Curve

## 4.2 Hasil dan Analisa Model Prediksi

### Performance metrik;

Metrik	Value
Accuracy	0,99999033
Precision	0,99997148
Recall	1
F1-Score	0,99998574

### Penjelasan:

Akurasi, model mencapai akurasi  $\approx 99,99\%$ , yang berarti model memprediksi hasil dengan tepat untuk hampir semua kasus. Presisi  $\approx 99,99\%$ , yang menunjukkan bahwa ketika model memprediksi tabrakan, hasilnya hampir selalu benar. Recall = 100%, yang berarti model berhasil mengidentifikasi semua kasus tabrakan yang sebenarnya. Skor F1  $\approx 99,99\%$ , yang menyeimbangkan presisi dan recall, menunjukkan efektivitas model secara keseluruhan.

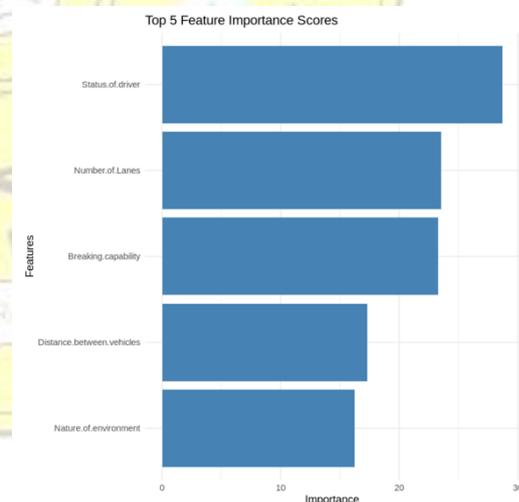
### 4.3 Analisa Fitur Importance

Fitur	Importance
Status.of.driver	28,6831548
Number.of.Lanes	23,5290305
Breaking.capability	23,2804059
Distance.between.vehicles	17,3155748
Nature.of.environment	16,2228242
Velocity.of.vehicle	2,38017622
Predicted_Collussion	0,4173755

### Penjelasan Pentingnya Fitur:

- 1) Status Pengemudi (*Status.of.driver*): Fitur ini memiliki skor kepentingan tertinggi, yang menunjukkan bahwa kondisi atau perilaku pengemudi merupakan faktor paling krusial dalam memprediksi tabrakan.
- 2) Jumlah Jalur (*Number.of.Lanes*): Fitur terpenting kedua, menunjukkan bahwa struktur jalan berdampak signifikan terhadap kemungkinan tabrakan.
- 3) Kemampuan Pengereman (*Breaking.capability*): Fitur ini berada di peringkat ketiga, yang menyoroti pentingnya kinerja kendaraan dalam pencegahan tabrakan.
- 4) Jarak Antar Kendaraan (*Distance.between.vehicles*): Fitur terpenting keempat, yang menekankan peran menjaga jarak aman dalam menghindari kecelakaan.
- 5) Sifat Lingkungan: Fitur kelima, yang menunjukkan bahwa faktor lingkungan (misalnya, cuaca, pencahayaan) juga memainkan peran penting.

### 4.3 Visualisasi Skore Fitur Importance



### Pembahasan;

- 1) Grafik batang di atas menunjukkan tingkat kepentingan relatif dari 5 fitur teratas dalam memprediksi tabrakan:
- 2) Status Pengemudi adalah fitur yang paling berpengaruh.

- 3) Jumlah Jalur dan Kemampuan Mengerem mengikuti dengan tingkat kepentingan yang sama.
- 4) Jarak Antar Kendaraan dan Kondisi Lingkungan juga berkontribusi secara signifikan.

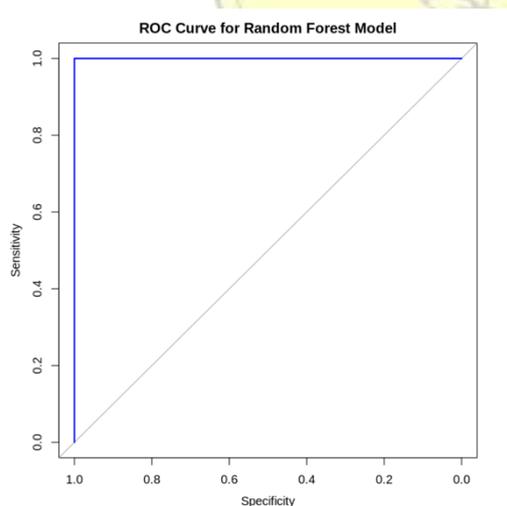
#### 4.4 Confusion Matrix & Kurva ROC

Confusion

**True Negatives (0,0):** Sebanyak 70.112 kasus berhasil diprediksi dengan benar sebagai bukan tabrakan. **True Positives (1,1):** Sebanyak 136.661 kasus berhasil diprediksi dengan benar sebagai tabrakan. **False Negatives (0,1):** Hanya terdapat 2 kasus yang salah di-klasifikasikan sebagai bukan tabrakan, yang sebenarnya terjadi tabrakan. **False Positives (1,0):** Tidak ada kasus yang salah diklasifikasikan sebagai tabrakan.

Predicted	0	1
0	70.112	2
1	0	136.616

Kurva ROC menunjukkan kemampuan model dalam membedakan antara dua kelas. AUC (*Area Under the Curve*): "AUC: 0,999999919690797", nilai ini menunjukkan dan mengindikasikan bahwa model memiliki kemampuan diskriminasi yang hampir sempurna dalam membedakan antara kasus tabrakan dan bukan tabrakan.



## 5 KESIMPULAN

Penelitian ini membuktikan bahwa pendekatan machine learning dalam konteks *Internet of Vehicles (IoV)* sangat efektif dalam membangun sistem prediksi tabrakan yang cerdas dan adaptif. Dengan memanfaatkan data dinamis seperti kecepatan kendaraan, jarak antar kendaraan, status pengemudi, dan kondisi lingkungan, model yang dikembangkan mampu membedakan kondisi berisiko dan tidak berisiko tabrakan secara akurat. Proses pembangunan model melalui tahapan sistematis—mulai dari eksplorasi data, preprocessing, seleksi fitur, pelatihan model, hingga evaluasi performa—telah menghasilkan kinerja prediktif yang sangat baik.

Model klasifikasi yang digunakan, seperti SVM, Random Forest, dan MLP, menunjukkan performa tinggi, terutama pada model yang menghasilkan **akurasi 99,99%**, **presisi 99,99%**, **recall 100%**, dan **F1-score 99,99%**. Hasil ini menegaskan bahwa model tidak hanya tepat dalam mendeteksi kejadian tabrakan, tetapi juga sangat sensitif terhadap semua kasus yang sebenarnya terjadi, tanpa kesalahan klasifikasi yang berarti. Bahkan, pada evaluasi *confusion matrix*, hanya dua kasus yang salah diklasifikasikan sebagai bukan tabrakan, dan tidak ada prediksi tabrakan yang salah.

Fitur yang paling berkontribusi terhadap performa model adalah **status pengemudi**, diikuti oleh **jumlah jalur**, **kapabilitas pengereman**, dan **jarak antar kendaraan**, yang semuanya masuk akal dalam konteks keselamatan berkendara. Visualisasi ROC curve yang menghasilkan AUC mendekati sempurna (0.9999999) memperkuat bukti bahwa model memiliki kemampuan diskriminasi yang luar biasa antara dua kelas prediksi.

Dengan demikian, model ini tidak hanya unggul secara teoritis, tetapi juga sangat potensial untuk diimplementasikan secara praktis dalam sistem peringatan dini tabrakan di kendaraan cerdas. Penelitian ini memberikan fondasi penting bagi pengembangan sistem keamanan lalu lintas

berbasis IoV yang responsif, presisi, dan berorientasi pada pencegahan kecelakaan.

### Daftar Pustaka

- Almutairi, M. S., Almutairi, K., & Chiroma, H. (2023a). Hybrid of deep recurrent network and long short term memory for rear-end collision detection in fog based internet of vehicles. *Expert Systems with Applications*, 213, 119033. <https://doi.org/10.1016/j.eswa.2022.119033>
- Almutairi, M. S., Almutairi, K., & Chiroma, H. (2023b). Selecting Features That Influence Vehicle Collisions in the Internet of Vehicles Based on a Multi-Objective Hybrid Bi-Directional NSGA-III. *Applied Sciences (Switzerland)*, 13(4). <https://doi.org/10.3390/app13042064>
- Chang, C. C., Ooi, Y. M., & Sieh, B. H. (2021). IoV-Based Collision Avoidance Architecture Using Machine Learning Prediction. *IEEE Access*, 9, 115497–115505. <https://doi.org/10.1109/ACCESS.2021.3105619>
- Jeong, Y., Kim, S., & Yi, K. (2020). Surround vehicle motion prediction using lstm-rnn for motion planning of autonomous vehicles at multi-lane turn intersections. *IEEE Open Journal of Intelligent Transportation Systems*, I(1), 2–14. <https://doi.org/10.1109/OJITS.2020.2965969>
- Manchala, A. K., & Vijaya Kishore, V. (2024). Advancements in Machine Learning and Data Mining Techniques for Collision Prediction and Hazard Detection in Internet of Vehicles. *Passer Journal of Basic and Applied Sciences*, 6(1), 80–91. <https://doi.org/10.24271/PSR.2023.403821.1342>
- Qian, S. (2022). Vehicle Collision Prediction Model on the Internet of Vehicles. *Lecture Notes in Electrical Engineering, 942 LNEE*, 518–530. [https://doi.org/10.1007/978-981-19-2456-9\\_53](https://doi.org/10.1007/978-981-19-2456-9_53)
- Strickland, M., Fainekos, G., & Amor, H. Ben. (2018). Deep predictive models for collision risk assessment in autonomous driving. *Proceedings - IEEE International Conference on Robotics and Automation*, 4685–4692. <https://doi.org/10.1109/ICRA.2018.8461160>
- Yang, L., Wang, Z., Ma, L., & Dai, W. (2022). Transfer Learning-Based Vehicle Collision Prediction. *Wireless Communications and Mobile Computing*, 2022. <https://doi.org/10.1155/2022/2545958>
- Chiroma Gombe, H. (2023). Internet of Vehicles Dataset [Data set]. Kaggle. <https://www.kaggle.com/datasets/harunachiromagombe/internet-of-vehicles-dataset>