

**PERBANDINGAN METODE RANDOM FOREST DAN
CNN DENGAN MODEL RESNET-50 PADA
KLASIFIKASI MIT-BIH ARRHYTHMIA DATABASE**

PROPOSAL PROGRAM INTERNSHIP I

Diajukan untuk memenuhi kelulusan mata kuliah Internship

Tahun akademik **2023/2024**

Disusun Oleh:

Nama : **M. RIZKY**

NPM : 1194021



Universitas Logistik & Bisnis Internasional

**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA
UNIVERSITAS LOGISTIK & BISNIS INTERNASIONAL
BANDUNG**

2023

***COMPARISON OF RANDOM FOREST AND CNN METHODS WITH
RESNET-50 MODELS IN THE MIT-BIH ARRHYTHMIA DATABASE
CLASSIFICATION***

INTERNSHIP PROGRAM REPORT I

Submitted to allow Internship course graduation

Academic year **2023/2024**

Created By:

Name : **M. RIZKY**

NPM : 1194021



Universitas Logistik & Bisnis Internasional

***APPLIED BACHELOR PROGRAM OF INFORMATICS ENGINEERING
UNIVERSITY OF INTERNATIONAL LOGISTICS & BUSINESS
BANDUNG***

2023

LEMBAR PENGAJUAN PEMBIMBING

PERBANDINGAN METODE RANDOM FOREST DAN CNN DENGAN MODEL RESNET-50 PADA KLASIFIKASI MIT-BIH ARRHYTHMIA DATABASE

M. RIZKY : 1.19.4.021

Proposal ini diajukan untuk permohonan pengajuan pembimbing *internship I*

Bandung, 10 Maret 2023

Pembimbing Eksternal,

Pembimbing Internal,

NIP : XXX.XXX

Roni Andarsyah, S.T., M.Kom., SFPC

NIP : XXX.XXX

Menyetujui,
Koordinator Program Internship I

Mohamad Nurkamal Fauzan, S.T., M.T., SFPC

NIP : XXX.XXX

ABSTRAK

Di era modern ini, banyak sekali kegiatan – kegiatan operasional maupun kegiatan lainnya yang melibatkan atau menggunakan *Artificial Intelligence (AI)*. Tidak bisa dipungkiri lagi bahwa teknologi yang berkembang pesat seperti sekarang ini tentu dibuat untuk mempermudah pekerjaan manusia bahkan menggantikan peran manusia terutama dibidang kesehatan.

Machine Learning sebagai metode dalam sistem kecerdasan buatan yang mampu mengklasifikasikan data yang dimasukkan untuk keperluan dan kebutuhan masing - masing. Banyak aplikasi yang menerapkan *Machine Learning* maupun *Deep Learning* untuk keperluan klasifikasi data, memprediksi hubungan antar data, membaca pola data, dan banyak implementasi lainnya. Sedangkan *Deep Learning*, merupakan kemajuan terbaru dari kecerdasan buatan yang telah digunakan mengatasi banyak masalah yang melibatkan kumpulan data besar atau big data

Dalam laporan ini akan dibahas perbandingan antara *Machine Learning* dan *Deep Learning* digunakan untuk mendeteksi sinyal EKG apakah sudah sesuai dengan data yang sudah di tetapkan pada anotasi. Sehingga akan didapatkan metode yang cocok dalam melakukan klasifikasi.

Kata Kunci: *Artificial Intelligence, Machine Learning, Deep Learning,*
Klasifikasi, EKG

ABSTRACT

In this modern era, there are many operational activities and other activities that involve or use Artificial Intelligence (AI). It cannot be denied that technology that is developing rapidly like today is certainly made to facilitate human work and even replace the role of humans, especially in the health sector.

Machine Learning as a method in an artificial intelligence system that is able to classify the data entered for their respective needs and requirements. Many applications implement Machine Learning and Deep Learning for data classification purposes, predicting relationships between data, reading data patterns, and many other implementations. While Deep Learning, is the latest advancement of artificial intelligence which has been used to overcome many problems involving large data sets or big data

This report will discuss a comparison between Machine Learning and Deep Learning which is used to detect whether the ECG signal is in accordance with the data specified in the annotation. So that a suitable method will be obtained in carrying out the classification.

Keywords: *Artificial Intelligence, Machine Learning, Deep Learning, Classification, ECG*

KATA PENGANTAR

Dalam rangka menyelesaikan program studi yang akan saya jalani, saya ingin mengajukan proposal internship 2 dengan judul “Perbandingan Metode Random Forest dan CNN dengan model ResNet-50 pada Klasifikasi MIT-BIH Arrhythmia Database”.

Penelitian ini bertujuan untuk melakukan perbandingan antara model tradisional atau *Machine Learning* dengan model *Deep Learning* atau *Convolutional Neural Network (CNNs)* dalam melakukan klasifikasi Arrhythmia MIT-BIH Database.

Saya yakin bahwa hasil penelitian ini dapat memberikan kontribusi yang bermanfaat dalam mengembangkan kebijakan publik di Indonesia, khususnya dalam bidang kesehatan. Selain itu, melalui program internship ini, saya berharap dapat memperoleh pengalaman dan pengetahuan yang lebih luas dalam bidang penelitian dan pengembangan teknologi informasi.

Saya mengucapkan terima kasih atas perhatian dan dukungan yang diberikan, dan saya juga berharap proposal ini dapat diterima untuk dilaksanakan.

Bandung, 11 Maret 2023

M. RIZKY

DAFTAR ISI

ABSTRAK	i
<i>ABSTRACT</i>	<i>ii</i>
KATA PENGANTAR	iii
DAFTAR ISI	iv
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	2
1.3. Tujuan	2
1.4. Manfaat	2
BAB II LANDASAN TEORI	3
2.1. MIT-BIH Arrhythmia Database	3
2.2. Elektrokardiogram (EKG)	3
2.3. Machine Learning	4
2.3.1. Supervised Learning	5
2.3.1.1. Random Forest	5
2.3.1.2. Naive Bayes	6
2.3.1.3. K-Nearest Neighbor	7
2.3.1.4. Support Vector Machine (SVM)	7
2.3.2. Unsupervised Learning	8
2.3.2.1. K-Means	9
2.4. Deep Learning	9
2.4.1. Convolutional Neural Network (CNN)	10
BAB III METODOLOGI PENELITIAN	11
3.1. Ruang Lingkup	11
3.2. Alur Metodologi Penelitian	11
3.3. Indikator Capaian Penelitian	12
3.3.1. Studi Literatur	13
3.3.2. Pengumpulan Data	13
3.3.3. Pra-Pemrosesan Data	14
3.3.4. Pemodelan	14
3.3.5. Evaluasi Model	14
BAB IV PENUTUP	15
4.1. Kesimpulan dan Saran	15

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Tidak bisa dipungkiri lagi bahwa perkembangan Artificial Intelligence begitu sangat cepat. Seiring berkembangnya teknologi yang sangat cepat, banyak sekali jenis - jenis teknologi yang bermunculan untuk membantu bahkan menggantikan pekerjaan manusia salah satunya di bidang kesehatan. Salah satu cabang Artificial Intelligence yang sekarang banyak sekali diminati adalah Machine Learning dan sampai saat ini masih terus berkembang pesat di kalangan programmer atau khususnya di dunia IT. Machine Learning sendiri terdiri atas 2 bagian yaitu Machine Learning Supervised dan Machine Learning Unsupervised. Sedangkan *Deep Learning*, merupakan kemajuan terbaru dari kecerdasan buatan yang telah digunakan mengatasi banyak masalah yang melibatkan kumpulan data besar atau big data.

Machine Learning Supervised adalah struktur dari suatu data yang hendak dianalisis telah ditentukan dahulu dan Machine Learning mencari data di struktur tersebut, sedangkan Machine Learning Unsupervised struktur dari suatu data dicari oleh Machine Learning itu sendiri [2]. Salah satu algoritma Supervised Learning yang sering digunakan pada proses klasifikasi adalah algoritma Random Forest (RF). RF adalah teknik bagging yang memiliki karakteristik signifikan yang berjalan efisien pada dataset besar. Random forest dapat menangani ribuan variabel masukan tanpa penghapusan variabel dan memperkirakan fitur penting untuk klasifikasi [1].

Di dalam dunia medis, teknologi - teknologi banyak sekali diterapkan untuk memenuhi kebutuhan medis itu sendiri seperti AI pendeteksi pasien positif Covid atau tidak dengan memanfaatkan hembusan nafas dari pasien tersebut. Dan masih banyak lagi hal - hal yang bisa kita manfaatkan untuk mengembangkan teknologi di bidang kesehatan salah satunya adalah dengan

memanfaatkan sinyal EKG atau Elektrokardiogram untuk mengklasifikasi penyakit gagal jantung pada pasien. EKG merupakan sebuah informasi sinyal yang digambarkan dalam bentuk diagram yang menampilkan informasi penting mengenai keadaan jantung manusia. Elektrokardiogram atau EKG adalah rekaman aktivitas listrik yang dihasilkan melalui siklus detak jantung [2].

1.2. Rumusan Masalah

1. Bagaimana proses ekstraksi fitur pada sinyal EKG?
2. Algoritma yang cocok untuk proses klasifikasi pada sinyal EKG?
3. Bagaimana proses implementasi model yang akan digunakan untuk mengklasifikasi data pada sinyal EKG?

1.3. Tujuan

Dari penjelasan pada bagian latar belakang di atas, dapat kita pahami bahwa penulisan laporan ini memiliki tujuan antara lain:

1. Dapat memahami bagaimana proses ekstraksi fitur pada sinyal EKG
2. Dapat menentukan model yang terbaik untuk proses klasifikasi sinyal EKG
3. Dapat memahami bagaimana implementasi model yang dipilih dalam proses klasifikasi

1.4. Manfaat

Manfaat yang bisa didapatkan pada penelitian ini adalah kita dapat memahami bagaimana proses dari kita melakukan ekstraksi ciri dari sinyal EKG lalu kemudian melakukan preprocessing data yang dimana proses tersebut harus kita lakukan agar data yang diproses klasifikasikan menghasilkan output yang akurat dengan hasil yang seharusnya.

BAB II

LANDASAN TEORI

2.1. *MIT-BIH Arrhythmia Database*

MIT-BIH Arrhythmia database adalah rangkaian uji standar yang umumnya tersedia untuk mengevaluasi aritmia deteksi. Sejak 1980, *database* ini telah digunakan untuk dasar penelitian untuk dinamika jantung di sekitar 500 lokasi di seluruh dunia. Basis data ini sebagian besar digunakan untuk tujuan medis dan penelitian dari deteksi dan analisis aritmia jantung yang berbeda. Basis data ini mencoba menyediakan informasi yang tepat untuk mendeteksi aritmia ventrikel [3].

Aritmia adalah perubahan detak jantung yang tidak normal karena detak jantung yang tidak tepat yang menyebabkan kegagalan dalam pemompaan darah. Aritmia dapat menyebabkan kematian jantung mendadak. Gejala aritmia yang umum adalah denyut prematur, jantung berdebar, pusing, kelelahan, dan pingsan. Aritmia lebih sering terjadi pada orang yang menderita tekanan darah tinggi, diabetes dan arteri koroner penyakit [4]. Sinyal *Electrocardiogram* yang akan digunakan pada kegiatan internship ini diambil dari *MIT-BIH Arrhythmia database*.

2.2. *Elektrokardiogram (EKG)*

Elektrokardiogram (EKG) adalah tes medis yang mengukur aktivitas listrik jantung. EKG digunakan untuk mendiagnosis dan memantau berbagai kondisi jantung, seperti serangan jantung, aritmia, dan gagal jantung. EKG akan merekam aktivitas listrik kecil yang dihasilkan oleh jantung selama periode waktu tertentu dengan menempatkan elektroda pada tubuh pasien [5]. Rekaman EKG berisi *noise* dan amplitudo yang bervariasi dari setiap orang sehingga sulit dalam proses mendiagnosis [6].

Elektrokardiogram (EKG) memberikan informasi penting tentang berbagai kondisi manusia [7]. Untuk melakukan EKG, petugas kesehatan akan menempelkan tambalan kecil dan lengket yang disebut elektroda ke dada, lengan, dan kaki pasien. Elektroda terhubung ke mesin EKG, yang merekam sinyal listrik yang dihasilkan oleh jantung saat bergerak ke seluruh tubuh. Mesin tersebut menghasilkan jejak aktivitas listrik jantung, yang disebut strip EKG, yang kemudian diinterpretasikan oleh petugas kesehatan.

Machine Learning dapat diterapkan pada analisis data EKG untuk meningkatkan akurasi dan efisiensi diagnosis dan pengobatan. Misalnya, algoritma pembelajaran mesin dapat dilatih untuk mengenali pola dalam data EKG yang menunjukkan kondisi jantung tertentu. Algoritma ini kemudian dapat digunakan untuk menganalisis data EKG secara otomatis dan memberikan rekomendasi diagnosis atau pengobatan. Algoritma pembelajaran mesin dapat dilatih untuk mengklasifikasikan data EKG ke dalam kategori yang berbeda, seperti normal atau abnormal, atau untuk mengidentifikasi kondisi jantung tertentu.

Biasanya, klasifikasi sinyal EKG memiliki empat fase: preprocessing, segmentasi, ekstraksi fitur, dan klasifikasi. Fase preprocessing terutama ditujukan untuk mendeteksi dan melemahkan frekuensi sinyal EKG yang terkait dengan artefak, yang juga biasanya melakukan normalisasi dan peningkatan sinyal. Setelah preprocessing, segmentasi akan membagi sinyal menjadi segmen yang lebih kecil, yang dapat mengekspresikan aktivitas listrik jantung dengan lebih baik [6].

2.3. Machine Learning

Pembelajaran Mesin atau Machine Learning merupakan kemajuan teknologi yang penting karena dapat membantu dalam mengambil keputusan dengan mekanisme prediksi dan klasifikasi berdasarkan data yang ada [8]. Berfokus pada performance yang tinggi, teknik pembelajaran mesin atau machine learning diterapkan pada bisnis dengan data yang berkembang pesat. Karena

pendekatan desain cocok untuk komunikasi komputasi paralel dan terdistribusi yang berevolusi atau data bisnis yang dinamis dan berkembang kedalam model Machine Learning [9].

Teknologi berbasis komputer modern banyak yang telah menggunakan Pembelajaran Mesin atau Machine Learning. Machine Learning merupakan cabang dari Kecerdasan Buatan atau Artificial Intelligence yang luas dan sudah berkembang pesat saat ini yang memungkinkan komputer untuk belajar dan berkembang secara otomatis tanpa harus diprogram secara eksplisit. Teknologi ini berasal dari mempelajari pengenalan pola dan teori pembelajaran komputasi. Secara umum, metode pembelajaran yang umum digunakan oleh Machine Learning dapat diklasifikasikan menjadi Supervised, Unsupervised, dan Reinforcement Learning [10].

2.3.1. Supervised Learning

Supervised Learning merupakan metode Machine Learning untuk menyimpulkan fungsi dari data *train* ada. Algoritma Supervised Learning biasanya berisi kumpulan sampel input (*feature*) dan label yang berkaitan dengan kumpulan data tersebut. Tujuan dari pengklasifikasian adalah untuk menemukan batas yang sesuai yang dapat memprediksi label yang benar pada data *test*. Secara singkat, dalam Supervised Learning memiliki setiap contoh data yang berpasangan yang terdiri dari objek masukan (*input*) dan objek keluaran (*output*) yang diinginkan. Algoritma Supervised Learning menganalisis data *train* dan menghasilkan fungsi (*model*) [11]. Beberapa contoh metode algoritma pada Supervised Learning:

2.3.1.1. Random Forest

Random Forest adalah metode Machine Learning yang diperkenalkan pada tahun 2001 oleh Leo Breiman. Metode ini menggunakan serangkaian besar dari Decision Tree dengan korelasi timbal balik yang rendah dan fitur yang dipilih secara acak menggunakan metode bagging (Bootstrap AGGREGatING) [12].

Random Forest merupakan salah satu metode pengklasifikasian terbaik dan banyak digunakan untuk regresi dan juga klasifikasi. Random Forest memiliki algoritma yang sederhana sehingga menjadi salah satu pilihan yang menarik untuk mengklasifikasi teks. Selain itu, Random Forest juga memiliki kemampuan untuk mengolah data berdimensi tinggi dan memiliki performa yang tinggi walaupun menggunakan data yang banyak sehingga menjadi salah satu keuntungan menggunakan model ini dibandingkan dengan model Machine Learning lainnya [13].

Pemilihan model ini didasarkan karena pada faktanya bahwa Random Forest secara luas dianggap sebagai salah satu metode Machine Learning yang paling sukses dan banyak digunakan hingga saat ini [14].

2.3.1.2. Naive Bayes

Naive Bayes adalah salah satu pengklasifikasi terkemuka yang telah banyak dikutip oleh banyak peneliti dan digunakan di banyak aspek karena kesederhanaannya dan kinerja dari klasifikasi yang nyata [15]. Diantara bermacam-macam teknik atau metode klasifikasi saat ini, pengklasifikasi Naive Bayes (*NB*) berperan penting karena kesederhanaan, traktabilitas dan efisiensinya [16].

Naive Bayes juga merupakan pengklasifikasi yang sangat kompeten dalam banyak aplikasi di dunia nyata. Meskipun Naive Bayes telah menunjukkan akurasi klasifikasi yang luar biasa, namun output yang dihasilkan jarang benar dalam kenyataan [17]. Terlepas dari hal itu, pada kenyataannya Naive Bayes bekerja dengan baik diimplementasikan di dunia ini seperti memprediksi waktu, pemfilteran spam, prakiraan cuaca, dan diagnosis medis [18].

Pengklasifikasian Naive Bayes didasarkan pada kombinasi Teorema Bayes dan asumsi independensi atribut. Pengklasifikasi Naive

Bayes didasarkan pada asumsi yang disederhanakan bahwa nilai atribut bersifat independen secara kondisional, berdasarkan asumsi nilai target yang diberikan. Pendekatan Bayes untuk klasifikasi kasus baru terdiri dari penetapan nilai target yang paling mungkin, dengan asumsi bahwa ada [19].

2.3.1.3. K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan salah satu algoritma klasifikasi yang paling stabil dalam kelompok algoritma klasifikasi supervised. Dikarenakan kesederhanaan dan implementasi algoritma yang mudah.

K-Nearest Neighbor (KNN) merupakan salah satu metode klasifikasi nonparametric. KNN menjadi terkenal karena algoritmanya yang luas dan yang paling mudah. KNN dapat menyimpan semua masalah atau studi kasus yang ada dan mengklasifikasikan berdasarkan kesamaannya. Secara umum, KNN menggunakan jarak Euclidean untuk menemukan data yang paling mirip dengan kelompoknya [20].

Pada metode ini, nilai yang hilang dari variabel tertentu diganti dengan nilai rata-rata atau nilai mean dari KNN terdekat dari pengamatan variabel yang sama. Fungsi jarak yang berbeda dapat digunakan untuk memilih tetangga yang memungkinkan metode untuk menyertakan variabel numerik dan kategori. Keuntungan utama KNN adalah tidak memerlukan spesifikasi model prediktif apapun [21].

2.3.1.4. Support Vector Machine (SVM)

Support Vector Machine membuktikan bahwa salah satu algoritma yang memiliki performance yang powerful selama beberapa dekade terakhir dan mengandalkan prinsip SRM. Support Vector Learning (SVM) umumnya digunakan untuk masalah klasifikasi dan regresi.

Support Vector Machine (SVM) bekerja dengan membangun hyperplane yang memisahkan sampel berdasarkan pendekatan margin yang maksimum. Berbeda dengan Artificial Neural Network (ANN) yang memiliki kelemahan local minimal. Support Vector Machine memberikan solusi dengan menyelesaikan masalah optimasi dengan konveks [22].

Metode Support Vector Machine juga disebut model klasifikasi biner. Dalam ruang dua dimensi, garis lurus menjadikan garis segmentasi yang paling cocok di tengah 2 kelas data, dan untuk kumpulan data berdimensi tinggi, ini untuk menetapkan bidang keputusan yang optimal sebagai tolak ukur klasifikasi. Prinsip dasar Support Vector Machine (SVM) mensyaratkan bahwa ketika masalah klasifikasi diselesaikan, jarak dari titik sampel terdekat ke permukaan keputusan adalah yang terbesar, yaitu jarak minimum memaksimalkan dua kelas titik sampel untuk memisahkan tepi [23].

2.3.2. Unsupervised Learning

Unsupervised Learning hanya dapat digunakan untuk tugas pengelompokan (clustering). Banyak pendekatan menggunakan Unsupervised Learning untuk mendukung tugas klasifikasi. Misalnya, algoritma pengelompokan (clustering) dapat meningkatkan kinerja tugas klasifikasi dengan mengelompokkan objek data ke dalam kelompok yang lebih homogen.

Unsupervised Learning banyak digunakan untuk preprocessing data seperti ekstraksi fitur, pemilihan fitur, dan resampling. Namun, ada banyak juga kasus penggunaan pembelajaran tanpa pengawasan sebagai algoritma pilihan untuk klasifikasi dengan kinerja yang sebanding dan mungkin lebih baik dibandingkan Supervised Learning [24].

Pada algoritma Unsupervised Learning yang mampu memisahkan data tanpa sebuah pengetahuan yang dalam tentang berbagai jenis peristiwa

meningkatkan efisiensi analisis secara luar biasa, dan memungkinkan analisis hilir untuk berkonsentrasi pada upaya penyesuaian hanya pada peristiwa yang menarik. Selain itu, algoritma pengelompokan memungkinkan lebih banyak eksplorasi data, berpotensi memungkinkan jenis reaksi baru dan tak terduga [25].

2.3.2.1. K-Means

Di era big data, sejumlah besar sumber daya data dikumpulkan dari kehidupan orang sehari-hari, ditransfer ke dalam internet, dan disimpan pada pusat data [26]. Pengelompokan data (Clustering), sebagai bagian penting dari data mining, dan sudah dianggap sebagai tugas penting dalam Unsupervised Learning.

Untuk kumpulan data tertentu, clustering akan membaginya menjadi beberapa kelompok atau cluster yang sedemikian rupa sehingga objek data dalam kelompok atau cluster yang sama berupa satu sama lain [27]. K-Means adalah pengelompokan masalah yang dipelajari dengan baik yang menghasilkan aplikasi di banyak bidang dan merupakan bagian dari Unsupervised Learning. K-Means merupakan salah satu masalah paling mendasar dalam ilmu komputer [28].

2.4. Deep Learning

Deep Learning, merupakan kemajuan terbaru dari kecerdasan buatan yang telah digunakan mengatasi banyak masalah yang melibatkan kumpulan data besar atau big data, Deep Learning juga muncul sebagai alat yang menjanjikan untuk analisis data scRNA-seq, karena memiliki kapasitas untuk mengekstrak fitur informatif dan ringkas data scRNA-seq yang berisik, heterogen, dan berdimensi tinggi untuk meningkatkan analisis hilir [29].

2.4.1. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs), telah berhasil digunakan untuk memproses dan menganalisa gambar digital, termasuk berbagai jenis gambar medis, termasuk CT. Meskipun beberapa karya telah mempresentasikan hasil menggunakan gambar sinar-X, CT scan cenderung lebih efisien karena dua alasan. Pertama, CT scan memberikan tampilan 3 dimensi yang mendetail dari organ yang didiagnosis, sedangkan sinar-X memberikan tampilan 2-D. Kedua, CT scan tidak tumpang tindih dengan organ, sedangkan pada sinar-X, tulang rusuk tumpang tindih dengan paru-paru dan jantung [30].

BAB III

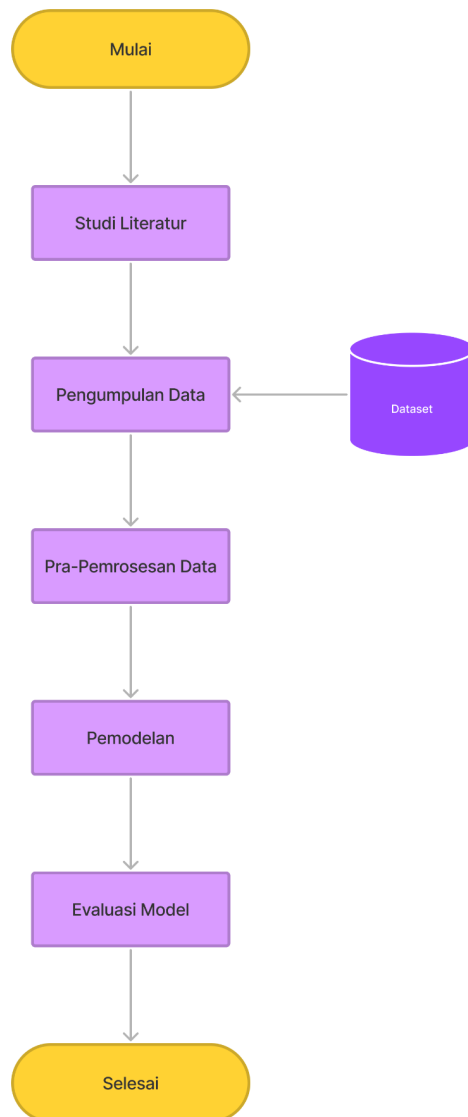
METODOLOGI PENELITIAN

3.1. Ruang Lingkup

Penelitian ini akan membahas tentang Perbandingan antara metode Random Forest dan CNN dengan model ResNet-50 pada klasifikasi arrhythmia database. Sumber data yang akan digunakan berasal dari situs PhysioNet yang merupakan Database Complex Physiologic Signals. Data yang akan diambil pada penelitian ini adalah MIT-BIH Arrhythmia Database dengan 48 record dan masing-masing durasi yang tersedia. Dari data-data tersebut akan digabungkan lalu dilakukan dengan masing-masing model.

3.2. Alur Metodologi Penelitian

Pada penelitian ini, metode akan menjadi hal penting dalam melakukan ekstraksi hingga klasifikasi sinyal karena akan mempengaruhi hasil klasifikasi atau output dari model tersebut. Oleh karena itu, alur metodologi yang tepat untuk mendapatkan hasil yang terbaik. Alur penelitian akan ditunjukan pada gambar 3.1.



Gambar 3.1 Flow Diagram Metodologi Penelitian

3.3. Indikator Capaian Penelitian

Berdasarkan gambar diatas, terdapat beberapa indikator capaian yaitu sebagai berikut:

No	Tahapan	Indikator Capaian
----	---------	-------------------

1	Studi Literatur	Uraian Teori - teori
2	Pengumpulan Data	Dataset sinyal MIT-BIH Arrhythmia
3	Pra-Pemrosesan Data	Ekstraksi Fitur, Denoising, Normalisasi
4	Pemodelan	Random Forest dan CNN dengan arsitektur ResNet - 50
5	Evaluasi Model	Melakukan perbandingan dengan beberapa model

Tabel 3.1 Indikator Capaian Penelitian

3.3.1. Studi Literatur

Pada studi literatur ini berisikan tentang uraian teori, temuan dan juga bahan penelitian orang lain yang didapatkan pada jurnal - jurnal nasional maupun internasional. Pada bagian ini akan dijadikan acuan dari kegiatan penelitian ini untuk mengimplementasikan dan juga mengembangkan dari beberapa penelitian yang ada.

3.3.2. Pengumpulan Data

Pada tahap ini, data yang akan digunakan adalah data yang diambil langsung dari website PhysioNet yang merupakan *Research Resource for Complex Physiological Signals* untuk melakukan penelitian dan pendidikan biomedis dan menawarkan akses gratis pada database yang disediakan. PhysioNet didirikan pada tahun 1999 dibawah naungan *National Institutes of Health (NIH)*.

Data yang dikumpulkan merupakan data hasil rekaman detak jantung dari beberapa orang. Data yang digunakan berbentuk CSV dengan nilai hasil dari indikator grafik yang digambarkan. Pada data ini juga

diberikan anotasi di setiap detak jantungnya untuk dijadikan label dari dataset tersebut.

3.3.3. Pra-Pemrosesan Data

Pra-Pemrosesan data merupakan proses pembersihan data dengan melakukan ekstraksi fitur dengan menggabungkan data menjadi satu data frame dengan masing-masing notasinya yang nantinya data tersebut akan dilakukan *denoising* agar data yang digunakan menjadi optimal dan tidak terdapat data-data yang *outlier* atau jauh dari nilai aslinya. Setelah itu, data yang sudah dilakukan denoising akan di normalisasi agar skala yang dipakai memiliki nilai yang sama.

3.3.4. Pemodelan

Pada tahap ini akan dilakukan pemodelan dengan menggunakan data yang sebelumnya. Data yang tersedia perlu dilakukan split untuk keperluan data *train* dan data *test* untuk nanti dilakukan proses pemodelan menggunakan data *train* tersebut.

3.3.5. Evaluasi Model

Tahap evaluasi model merupakan tahapan untuk mengukur kinerja dari model yang dihasilkan apakah sudah memiliki akurasi yang baik atau tidak sehingga pada proses ini kita bisa melakukan perbandingan kinerja antara model yang tersedia dan model mana yang memiliki akurasi yang baik sesuai dengan hasil pra-pemrosesan data tersebut.

BAB IV

PENUTUP

4.1. Kesimpulan dan Saran

Berdasarkan hasil kajian yang dibahas di atas, pada proposal ini dapat ditarik kesimpulan bahwa pada penelitian ini memiliki kesimpulan adalah sebagai berikut:

1. Melakukan perbandingan antara Model Tradisional atau *Machine Learning* dengan *Deep Learning* atau *Convolutional Neural Networks (CNNs)*.
2. Melakukan evaluasi model menggunakan *confusion matrix*.

DAFTAR PUSTAKA

- [1] Wasimuddin, M., Elleithy, K., Abuzneid, A. S., Faezipour, M., & Abuzaghleh, O. (2020). Stages-based ECG signal analysis from traditional signal processing to machine learning approaches: A survey. *IEEE Access*, 8, 177782-177803.
- [2] Pojon, M. (2017). *Using machine learning to predict student performance* (Master's thesis).
- [3] Kuila, S., Dhanda, N., & Joardar, S. (2020). Feature Extraction and Classification of MIT-BIH Arrhythmia Database. In *Proceedings of the 2nd International Conference on Communication, Devices and Computing* (pp. 417-427). Springer, Singapore.
- [4] Apandi, Z. F. M., Ikeura, R., & Hayakawa, S. (2018, August). Arrhythmia detection using MIT-BIH dataset: A review. In *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)* (pp. 1-5). IEEE.
- [5] Li, T., & Zhou, M. (2016). ECG classification using wavelet packet entropy and random forests. *Entropy*, 18(8), 285.
- [6] Kumar, R. G., & Kumaraswamy, Y. S. (2012). Investigating cardiac arrhythmia in ECG using random forest classification. *International Journal of Computer Applications*, 37(4), 31-34.
- [7] Wasimuddin, M., Elleithy, K., Abuzneid, A. S., Faezipour, M., & Abuzaghleh, O. (2020). Stages-based ECG signal analysis from traditional signal processing to machine learning approaches: A survey. *IEEE Access*, 8, 177782-177803.
- [8] Kamila, N. K., Frnda, J., Pani, S. K., Das, R., Islam, S. M., Bharti, P. K., & Muduli, K. (2022). Machine learning model design for high performance cloud computing & load balancing resiliency: An innovative approach. *Journal of King Saud University-Computer and Information Sciences*.
- [9] Renggli, C., Ashkboos, S., Aghagolzadeh, M., Alistarh, D., & Hoefler, T. (2019, November). Sparcml: High-performance sparse communication for machine learning. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 1-15).
- [10] Bhatt, H., Shah, V., Shah, K., Shah, R., & Shah, M. (2022). State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review. *Intelligent Medicine*.

- [11] Khanzadeh, M., Chowdhury, S., Marufuzzaman, M., Tschopp, M. A., & Bian, L. (2018). Porosity prediction: Supervised-learning of thermal history for direct laser deposition. *Journal of manufacturing systems*, 47, 69-82.
- [12] Khanzadeh, M., Chowdhury, S., Marufuzzaman, M., Tschopp, M. A., & Bian, L. (2018). Porosity prediction: Supervised-learning of thermal history for direct laser deposition. *Journal of manufacturing systems*, 47, 69-82.
- [13] Jalal, N., Mehmood, A., Choi, G. S., & Ashraf, I. (2022). A novel improved random forest for text classification using feature ranking and optimal number of trees. *Journal of King Saud University-Computer and Information Sciences*.
- [14] Makariou, D., Barrieu, P., & Chen, Y. (2021). A random forest based approach for predicting spreads in the primary catastrophe bond market. *Insurance: Mathematics and Economics*, 101, 140-162.
- [15] Alizadeh, S. H., Hediehloo, A., & Harzevili, N. S. (2021). Multi independent latent component extension of naive bayes classifier. *Knowledge-Based Systems*, 213, 106646.
- [16] Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., & Sillero-Denamiel, M. R. (2021). Variable selection for Naïve Bayes classification. *Computers & Operations Research*, 135, 105456.
- [17] Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems*, 192, 105361.
- [18] Shaban, W. M., Rabie, A. H., Saleh, A. I., & Abo-Elsoud, M. A. (2021). Accurate detection of COVID-19 patients based on distance biased Naïve Bayes (DBNB) classification strategy. *Pattern Recognition*, 119, 108110.
- [19] Andrejiova, M., & Grincova, A. (2018). Classification of impact damage on a rubber-textile conveyor belt using Naïve-Bayes methodology. *Wear*, 414, 59-67.
- [20] Zamri, N., Pairan, M. A., Azman, W. N. A. W., Abas, S. S., Abdullah, L., Naim, S., ... & Gao, M. (2022). River quality classification using different distances in k-nearest neighbors algorithm. *Procedia Computer Science*, 204, 180-186.
- [21] Cubillos, M., Wöhlk, S., & Wulff, J. N. (2022). A bi-objective k-nearest-neighbors-based imputation method for multilevel data. *Expert Systems with Applications*, 117298.
- [22] Kumar, B., & Gupta, D. (2021). Universum based Lagrangian twin bounded support vector machine to classify EEG signals. *Computer Methods*

and Programs in Biomedicine, 208, 106244.

[23] Zhang, H., Shi, Y., Yang, X., & Zhou, R. (2021). A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance. *Research in International Business and Finance*, 58, 101482.

[24] Jafari-Marandi, R. (2021). Supervised or unsupervised learning? Investigating the role of pattern recognition assumptions in the success of binary predictive prescriptions. *Neurocomputing*, 434, 165-193.

[25] Solli, R., Bazin, D., Hjorth-Jensen, M., Kuchera, M. P., & Strauss, R. R. (2021). Unsupervised learning for identifying events in active target experiments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1010, 165461.

[26] Zhao, D., Hu, X., Xiong, S., Tian, J., Xiang, J., Zhou, J., & Li, H. (2021). K-means clustering and kNN classification based on negative databases. *Applied Soft Computing*, 110, 107732.

[27] Wang, X., Wang, Z., Sheng, M., Li, Q., & Sheng, W. (2021). An adaptive and opposite K-means operation based memetic algorithm for data clustering. *Neurocomputing*, 437, 131-142.

[28] Zhang, Z., Feng, Q., Huang, J., Guo, Y., Xu, J., & Wang, J. (2021). A local search algorithm for k-means with outliers. *Neurocomputing*, 450, 230-241.

[29] Brendel, M., Su, C., Bai, Z., Zhang, H., Elemento, O., & Wang, F. (2023). Application of Deep Learning on Single-cell RNA Sequencing Data Analysis: A Review. *Genomics, Proteomics and Bioinformatics*. Beijing Genomics Institute.

[30] Naeem, H., & Bin-Salem, A. A. (2021). A CNN-LSTM network with multi-level feature extraction-based approach for automated detection of coronavirus from CT scan and X-ray images. *Applied Soft Computing*, 113.