

Pemodelan Hybrid Radial Basis Function Neural Network Berbasis Ensemble Learning dalam Prediksi Parameter Nitrat (NO_3^-) pada Sungai di Wilayah Tropis

Rifka Noor Azizah¹, Azahra Sukesih¹, Isti Dina Febriyanti¹, Nur Faizaturrohmah^{1*}, Fuji Lestari².

¹Program Studi Teknik Lingkungan, Institut Teknologi Sumatera, Lampung, Indonesia

²Program Studi Sains Aktuaria, Institut Teknologi Sumatera, Lampung, Indonesia

Correspondent Author: Nur Faizaturrohmah (nur.faizaturrohmah@tl.itera.ac.id)

ABSTRACT

Nitrate (NO_3^-) concentration is one of the key indicators of river water quality, particularly in tropical regions where high temperatures accelerate the nitrification process. Conventional methods such as STORET, the Pollution Index, and WQI/NSF can only assess actual conditions and are unable to project or provide early warning of water quality changes. This study aims to analyze the performance of a Hybrid Radial Basis Function Neural Network (Hybrid RBFNN) model in predicting nitrate concentration based on riverbank characteristics in tropical regions, while also identifying the water quality parameters that most influence nitrate levels through sensitivity analysis. The study was conducted using a systematic review approach on 40 selected publications (228 observation data points), which were subsequently categorized into residential riverbanks (99 data points) and agro-aquaculture riverbanks (129 data points). The Hybrid RBFNN model was developed by combining RBF, dense, and linear branches, and was further integrated with Gradient Boosting and Random Forest through a weighted-averaging ensemble approach. The results show that the model predicted nitrate concentration with high accuracy, achieving an R^2 value of 0.9909 for residential riverbanks and 0.9954 for agro-aquaculture riverbanks. Sensitivity analysis revealed that nitrite was the most influential parameter (40.25%), followed by electrical conductivity (19.80%) and TDS (11.10%). These findings indicate that the Hybrid RBFNN effectively models nonlinear relationships among water quality parameters and has the potential to be developed as a decision-support tool for an early warning system for nitrate pollution in tropical rivers.

Article History

Received 2026-05-01

Revised 2026-06-06

Accepted 2026-07-08

Keywords

radial basis function;
nitrate;
tropical rivers;
water quality;
neural network.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Pendahuluan

Sungai memiliki peranan penting dalam kehidupan sehari-hari, sungai memiliki peranan yang cukup penting. Kebutuhan akan air sungai dalam jumlah yang cukup dengan kualitas yang tinggi sering terjadi, ini menjadi kendala yang sering dihadapi terutama pada kawasan dengan tingkat aktifitas manusia yang tinggi. Sungai mengalir dari hulu dikawasan pegunungan menuju hilir yang berada di laut [1], [2]. Namun, peningkatan aktivitas manusia seperti pembuangan limbah industri, pertanian intensif, serta urbanisasi penduduk yang tidak terkendali di bantaran aliran sungai menyebabkan penurunan yang signifikan dalam kualitas

air [1], [3]. Jumlah air Sungai yang tersedia juga dapat dipengaruhi oleh perubahan iklim. Karena perubahan pola hujan dan peningkatan suhu dapat mengubah debit aliran, konsentrasi zat pencemar serta proses daur hidup yang ada di dalam perairan. Kondisi ini sering terjadi dan ditemukan pada sungai-sungai di wilayah tropis [4], [5]. Suhu tinggi yang mendorong dan mempercepat proses nitrifikasi di daerah tropis menyebabkan kadar nitrat (NO_3^-) meningkat dalam air sungai. Kadar nitrat (NO_3^-) adalah salah satu parameter pencemar utama yang digunakan untuk mengevaluasi kualitas air sungai [6]. Salah satu bentuk nitrogen utama di perairan alami adalah nitrat (NO_3^-), yang berasal dari ammonium hasil limbah yang masuk ke badan air. Aktivitas mikroorganisme mengoksidasi ammonium secara bertahap menjadi nitrit sebelum akhirnya membentuk nitrat (NO_3^-) di perairan [7].

Sebaliknya, metode seperti STORET, Indeks Pencemaran (IP), atau Indeks Kualitas Air (WQI/NSF) adalah yang biasa digunakan untuk mengukur status mutu air sungai, tetapi mereka hanya dapat memberikan penilaian kondisi aktual berdasarkan data laboratorium [8], [9]. Indeks tersebut tidak mampu memproyeksikan, ataupun memberikan peringatan dini terhadap perubahan kualitas air. Menurut Romdania et al [10], pendekatan deterministik seperti Storet dan IP sangat bergantung pada batas baku mutu dan tidak cukup responsif terhadap dinamika parameter yang berubah cepat seperti nitrat (NO_3^-).

Dalam sepuluh tahun terakhir, metode berbasis machine learning (ML) dan Artificial Neural Network (ANN) menjadi solusi inovatif dalam memprediksi kualitas air. ANN melakukannya dengan meniru cara kerja jaringan saraf biologis, dengan mampu membangun hubungan non-linear antara variabel input seperti pH, suhu, DO, salinitas, dan ion kimia lain dengan output misalnya TDS, nitrit, ammonia. Studi sebelumnya juga dilakukan oleh Gholamreza et al [11], menunjukkan bahwa Artificial Neural Network (ANN), terutama Fungsi Basis Radial (RBF) dan Time Delay Neural Network (TDNN), sangat efektif dalam memprediksi parameter Total Dissolved Solids (TDS) di Sungai Zayanderud di Iran. Dalam penelitian ini, data kualitas air yang dikumpulkan selama sepuluh tahun (2001–2010) digunakan. Variabel yang dimasukkan termasuk pH, Na^+ , Mg^{2+} , Ca^{2+} , HCO_3^- , Cl^- , dan total hardness (TH). Hasilnya menunjukkan bahwa model ANN mampu mencapai akurasi prediksi yang sangat tinggi, dengan koefisien determinasi (R^2) sebesar 0,997 dan indeks kesesuaian (IA) sebesar 0,999 pada jaringan RBF, sedangkan jaringan TDNN mencapai $R^2 = 0,957$ dan $\text{IA} = 0,986$. Temuan ini membuktikan bahwa model ANN lebih efisien dan andal dibandingkan metode statistik klasik dalam memprediksi perubahan kualitas air, khususnya nitrat (NO_3^-) yang sering kali dipengaruhi oleh faktor hidrologis dan kimia yang kompleks dan nonlinier.

Meskipun begitu, masih ada sejumlah kekurangan dalam penelitian yang belum dijawab oleh kajian-kajian sebelumnya. Pertama, mengenai perbandingan antar model ML, Li et al [12] melakukan analisis menyeluruh antara BPNN (Backpropagation Neural Network), RBFNN (Radial Basis Function Neural Network), SVM (Support Vector Machine), dan LSSVM (Least Squares Support Vector Machine) dalam memprediksi faktor kualitas air dalam budidaya, termasuk kadar nitrat (NO_3^-) dan nitrit (NO_2^-). Temuan dari studi ini mengungkapkan bahwa untuk perkiraan NO_3^- , SVM mencatat koefisien korelasi 0,99 Li et al [12], sedangkan RBFNN hanya mencapai 0,87 Li et al [12], menunjukkan bahwa RBFNN memiliki kekurangan signifikan dalam hal akurasi dibandingkan SVM, terutama dalam memperkirakan parameter nitrogen di dalam sistem perairan. Selain itu, dalam memprediksi NO_2^- , RBFNN menunjukkan koefisien korelasi serendah 0,08 Li et al [12], yang menunjukkan bahwa RBFNN cenderung menghasilkan kinerja yang sangat buruk pada parameter tertentu, tergantung pada karakteristik distribusi data input. Akan tetapi, penelitian oleh Li et al [12] dilakukan dalam lingkungan akuakultur yang terorganisir dan belum memperhitungkan variasi karakteristik tepi sungai di daerah tropis yang jauh lebih rumit dan beragam.

Lalu ada pula seperti LSTM yang memiliki kelebihan tersendiri untuk data deret waktu sungai Dodig et al [13], misalnya, memakai LSTM untuk memprediksi oksigen terlarut di

Sungai Sava, Serbia, dan hasilnya cukup tinggi (R^2 hingga 0,9998). Ini menunjukkan LSTM mampu menangkap pola perubahan kualitas air dari waktu ke waktu, terutama untuk data musiman, kemampuan yang tidak dimiliki RBFNN karena sifatnya statis. Namun penelitian tersebut tidak menyoroti parameter nitrat secara khusus, dan tidak mempertimbangkan perbedaan karakteristik bantaran sungai.

Dibanding model sekuensial, RBFNN lebih unggul saat data yang tersedia bukan berupa deret waktu panjang, melainkan data hasil kompilasi dari berbagai sumber seperti pada penelitian ini. RBFNN bisa langsung memodelkan hubungan nonlinier antarparameter tanpa perlu urutan waktu yang runtut, sehingga lebih cocok dipakai di sini. Hal ini juga sejalan dengan temuan Gholamreza et al [11], yang menunjukkan RBFNN unggul dibanding TDNN (R^2 0,997 vs 0,957) dalam memprediksi TDS di Sungai Zayanderud, Iran. Selain itu, arsitektur RBFNN yang lebih sederhana membuatnya lebih cepat dilatih dan lebih ringan dibanding model deep learning sekuensial, cocok untuk dataset dengan jumlah observasi yang terbatas.

Ada tiga hal yang jadi kebaruan dari penelitian ini. Pertama, penelitian ini secara khusus memprediksi nitrat menggunakan Hybrid RBFNN yang dipadukan dengan Gradient Boosting dan Random Forest lewat ensemble learning, pendekatan yang belum pernah dipakai pada studi RBFNN maupun LSTM sebelumnya. Kedua, penelitian ini membandingkan performa model pada dua karakteristik bantaran sungai sekaligus (pemukiman dan agro-akuakultur), untuk melihat apakah jenis aktivitas manusia di sekitar sungai berpengaruh pada akurasi prediksi. Ketiga, penelitian ini memakai analisis sensitivitas untuk mencari tahu parameter mana yang paling berpengaruh terhadap nitrat, sehingga ke depannya pemantauan bisa lebih efisien karena cukup mengukur parameter yang paling sensitif saja, tidak perlu mengukur semua parameter sekaligus. Model matematis konversi dari temuan ini memang belum bisa diwujudkan di penelitian ini karena datanya belum cukup lengkap, tapi identifikasi parameter sensitif ini jadi dasar penting sebelum model konversi tersebut bisa dibangun secara akurat di penelitian selanjutnya.

Tujuan dari penelitian ini adalah mengetahui kinerja model dari RBFNN berdasarkan karakteristik bantaran sungai pada wilayah tropis dengan menganalisis lebih lanjut parameter sensitivitas dari perbedaan karakteristik bantaran sungai dalam memprediksi kadar Nitrat (NO_3^-).

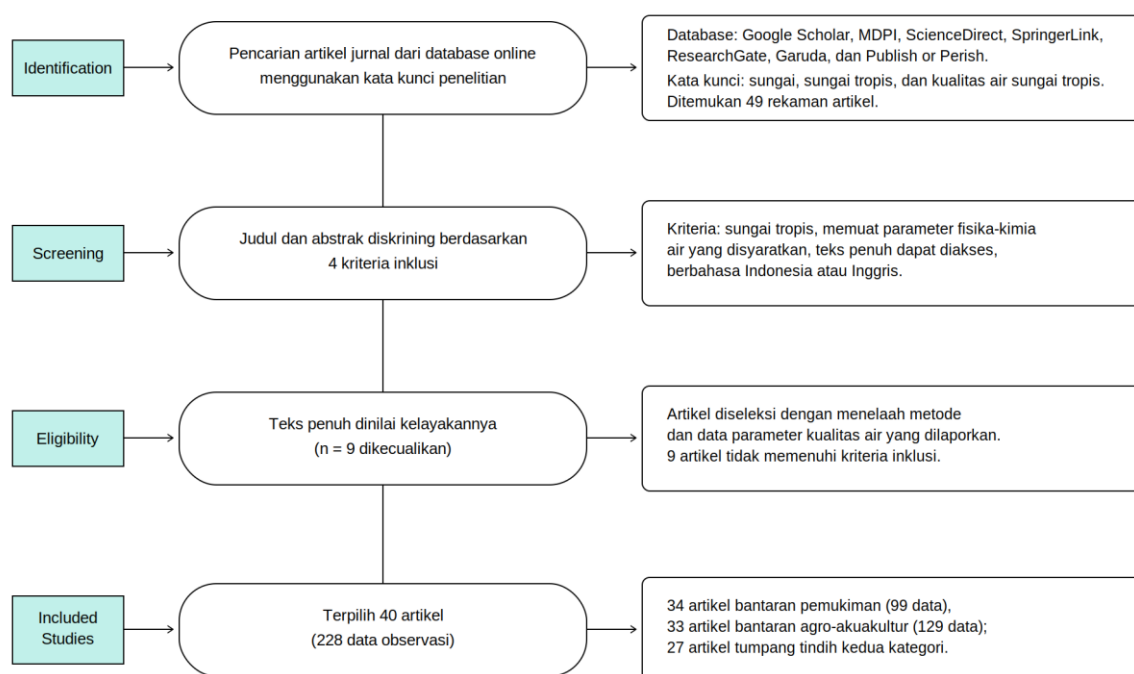
Metode

Penentuan Kriteria Review Literatur

Penelitian ini menggunakan pendekatan systematic review, yaitu metode tinjauan literatur yang dilakukan secara terstruktur dan mengikuti tahapan yang jelas mulai dari mengidentifikasi artikel, menyaring judul dan abstrak, menilai kelayakan teks penuh, hingga menentukan artikel yang benar-benar layak dianalisis. Studi ini menggunakan metode Hybrid Radial Basis Function Neural Network (Hybrid RBFNN) untuk memprediksi kualitas air sungai. Untuk melakukan ini, data dari berbagai sumber bacaan ilmiah digabungkan. Data yang digunakan dalam penelitian ini diperoleh dari berbagai repositori data akademik, seperti Google Scholar, MDPI, ScienceDirect, SpringerLink, ResearchGate, Garuda, dan Publish or Perish aplikasi Windows. Frasa kunci seperti "sungai", "sungai tropis", dan "kualitas air sungai tropis" digunakan untuk melakukan pencarian literatur. Skripsi, tugas akhir, dan laporan praktik kerja yang mengandung data uji mutu air sungai juga termasuk dalam kumpulan data atau informasi yang diperoleh. Proses penyaringan (screening) publikasi dilakukan berdasarkan empat kriteria inklusi, yaitu: 1) kajian dilakukan pada sungai di wilayah tropis; 2) publikasi memuat data parameter fisika-kimia mutu air, seperti suhu ($^{\circ}\text{C}$), TDS, pH, nitrat (NO_3^-), nitrit (NO_2^-), TSS, BOD, COD, DO, amonia (NH_3), dan DHL (daya hantar listrik); 3) naskah dapat diakses secara utuh (full text); dan 4) ditulis dalam bahasa Indonesia atau bahasa Inggris; 4) Terbitan publikasi yang digunakan dalam rentang 20 tahun terakhir untuk memastikan relevansi data dengan kondisi kualitas air sungai tropis. Berdasarkan penerapan keempat kriteria inklusi tersebut, diperoleh 40 terbitan yang memenuhi syarat dan digunakan sebagai sumber data dalam penelitian ini. Sebaliknya, artikel yang tidak memenuhi kriteria

tersebut dikeluarkan dari proses analisis (kriteria eksklusi). Mekanisme fisika dan kimia air yang dapat mempengaruhi kadar nitrat (NO_3^-) di sungai menentukan pilihan unsur-unsur ini. Untuk memudahkan pengolahan data, informasi dari terbitan yang memenuhi syarat selanjutnya dirangkum dalam format lembar kerja elektronik (.xlsx).

Menurut jenis penggunaan lahan di sepanjang tepi sungai, informasi dalam dokumen kerja ini dikategorikan menjadi dua kategori: 1) Tepi Sungai Permukiman: sungai yang mengalir melalui atau dekat dengan area tempat tinggal masyarakat di mana sumber utama pencemaran adalah air limbah rumah tangga, air limbah domestik, dan aliran air dari pemukiman di kota dan desa ; 2) Tepi Sungai Agro-Akuakultur, yaitu sungai yang berada di dekat areal yang didominasi oleh kegiatan pertanian, perkebunan, budidaya perairan (akuakultur), serta peternakan, di mana kontribusi nutrisi ke dalam badan air umumnya berasal dari pupuk berbasis nitrogen, pestisida, pakan ikan, dan limbah dari hewan ternak. Klasifikasi data ke dalam dua kelompok ini bertujuan untuk memahami apakah perbedaan kondisi lingkungan sekitar sungai, khususnya, jenis aktivitas manusia di tepi sungai dapat memengaruhi kemampuan model RBFNN untuk mengevaluasi hubungan antara parameter kualitas air dan kandungan nitrat (NO_3^-) sebagai variabel keluaran. Dengan cara ini, analisis sensitivitas model dapat dilakukan secara terpisah untuk setiap kelompok data. Ini memungkinkan pemahaman yang lebih baik tentang variabel input yang paling berdampak pada setiap jenis tepi sungai. Penelusuran literatur pada penelitian ini dilakukan dengan pendekatan systematic review, yaitu metode tinjauan pustaka yang dilakukan secara terstruktur dan mengikuti tahapan yang baku. Untuk memandu proses tersebut, digunakan kerangka PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), yaitu pedoman pelaporan yang lazim dipakai dalam systematic review untuk memastikan proses seleksi artikel berjalan transparan dan dapat ditelusuri ulang, mulai dari identifikasi rekaman artikel dari database, penyaringan judul dan abstrak, penilaian kelayakan teks penuh, hingga penentuan artikel yang akhirnya digunakan sebagai sumber data. Alur seleksi tersebut disajikan pada Gambar 1.



Gambar 1. Diagram Alur PRISMA Proses Seleksi Artikel

Perumusan Teknis Hybrid Radial Basis Function Neural Network (Hybrid RBFNN)

Model prediksi konsentrasi nitrat pada penelitian ini dikembangkan menggunakan Hybrid Radial Basis Function Neural Network (Hybrid RBFNN) yang diimplementasikan menggunakan bahasa pemrograman Python. Pengembangan model dilakukan dengan memanfaatkan pustaka TensorFlow-Keras untuk membangun arsitektur jaringan saraf, Scikit-learn untuk tahap praproses data, validasi model, dan implementasi algoritma pembandingan, serta NumPy, Pandas, dan Matplotlib untuk pengolahan data dan visualisasi hasil. Seluruh proses komputasi dijalankan pada lingkungan Google Colaboratory. Perangkat keras yang digunakan mengikuti spesifikasi komputasi yang disediakan oleh Google Colaboratory.

Secara konseptual, model ini mengadopsi struktur dasar Hybrid Radial Basis Function Neural Network (Hybrid RBFNN) yang terdiri atas input layer, hidden layer, dan output layer sebagaimana dijelaskan oleh Haykin [14]. Lapisan masukan menerima sepuluh parameter kualitas air, yaitu suhu, pH, TDS, TSS, DO, BOD, COD, nitrit (NO_2^-), amonia (NH_3), dan DHL (Daya Hantar Listrik), sedangkan lapisan keluaran menghasilkan prediksi konsentrasi nitrat (NO_3^-). Pada implementasinya, arsitektur tersebut dikembangkan menjadi Hybrid RBFNN melalui penambahan Dense Branch dan Linear Branch sehingga model mampu memanfaatkan representasi lokal dari fungsi radial sekaligus mempelajari hubungan nonlinier dan linier antarvariabel.

Jumlah neuron pada hidden layer tidak ditetapkan secara tetap, tetapi dihitung secara adaptif melalui konfigurasi implementasi program dengan jumlah maksimum 40 neuron. Penentuan jumlah neuron tersebut merupakan konfigurasi implementasi penelitian yang disesuaikan dengan ukuran dataset dan bukan merupakan ketentuan baku dalam metode RBFNN. Pada lapisan radial digunakan Gaussian Radial Basis Function sebagai fungsi aktivasi yang secara matematis dinyatakan sebagai:

$$\phi(x) = \exp(-\beta \|x - c\|^2) \quad (1)$$

dengan c menyatakan pusat (center) fungsi radial dan β menyatakan parameter penyebaran (spread parameter). Fungsi Gaussian merupakan fungsi aktivasi yang paling umum digunakan pada RBFNN karena mampu membentuk respons lokal terhadap data masukan dan memiliki kemampuan aproksimasi yang baik pada hubungan nonlinier (Bioshop [15]; Park & Sandberg, [16]; Haykin, [14]). Pada penelitian ini nilai awal parameter β diinisialisasi sebesar 1,0 dan ditetapkan sebagai parameter yang dapat diperbarui (trainable parameter), sehingga nilainya dioptimasi secara otomatis selama proses pelatihan. Sementara itu, pusat fungsi radial diinisialisasi menggunakan algoritma K-Means Clustering, yang merupakan pendekatan umum untuk memperoleh distribusi pusat neuron yang lebih representatif dibandingkan inisialisasi acak Moody & Darken, [17].

Arsitektur Hybrid RBFNN terdiri atas tiga cabang pemrosesan. Cabang pertama merupakan RBF Branch yang memanfaatkan keluaran neuron radial dan meneruskannya ke dua lapisan fully connected. Cabang kedua merupakan Dense Branch yang secara langsung mempelajari hubungan nonlinier dari data masukan menggunakan beberapa lapisan dense dengan fungsi aktivasi Rectified Linear Unit (ReLU). Untuk meningkatkan stabilitas pelatihan dan mengurangi kecenderungan overfitting, pada cabang ini diterapkan Batch Normalization dan Dropout sebesar 0,05. Cabang ketiga merupakan Linear Branch yang digunakan untuk mempertahankan informasi linier dari data masukan. Seluruh keluaran ketiga cabang kemudian digabungkan menggunakan operasi concatenate sebelum diproses pada lapisan keluaran dengan fungsi aktivasi linear, yang sesuai untuk permasalahan regresi kontinu [18].

Proses pelatihan model dilakukan menggunakan optimizer Adam dengan learning rate awal sebesar 0,001 dan Huber Loss sebagai fungsi kerugian. Optimizer Adam dipilih karena mengombinasikan estimasi momen orde pertama dan kedua sehingga banyak digunakan pada pelatihan jaringan saraf modern Kingma & Ba [19] sedangkan Huber Loss digunakan karena lebih stabil terhadap keberadaan pencilan dibandingkan Mean Squared Error pada permasalahan regresi Huber [7]. Untuk mencegah pelatihan berlebihan (overtraining),

diterapkan mekanisme Early Stopping dengan patience 150 epoch, serta Reduce Learning Rate on Plateau dengan factor 0,3, patience 50 epoch, dan minimum learning rate sebesar 1×10^{-8} . Setelah proses pelatihan utama selesai, dilakukan tahap fine tuning menggunakan learning rate sebesar 1×10^{-5} dengan Early Stopping sebesar 80 epoch.

Jumlah epoch maksimum pada tahap pelatihan utama ditetapkan sebanyak 5000 epoch, sedangkan pada tahap fine tuning ditetapkan maksimum 1000 epoch. Nilai batch size tidak ditentukan secara konstan, tetapi dihitung secara adaptif melalui konfigurasi program berdasarkan jumlah data pelatihan pada setiap tahap. Pendekatan ini merupakan bagian dari konfigurasi implementasi penelitian dan digunakan untuk menyesuaikan proses pembelajaran terhadap ukuran dataset yang digunakan.

Pendekatan Ensemble Learning

Untuk meningkatkan kemampuan prediksi, penelitian ini tidak hanya menggunakan Hybrid RBFNN sebagai model tunggal, tetapi menerapkan pendekatan ensemble dengan mengombinasikan tiga model, yaitu Gradient Boosting (GB), Random Forest (RF), dan Hybrid RBFNN. Ketiga model dilatih secara independen menggunakan dataset pelatihan yang sama sehingga masing-masing menghasilkan prediksi terhadap data uji.

Prediksi yang dihasilkan oleh ketiga model selanjutnya digabungkan menggunakan metode weighted averaging ensemble, yaitu kombinasi linier dari keluaran setiap model dengan bobot tertentu. Bobot masing-masing model tidak ditentukan secara subjektif, tetapi diperoleh melalui proses grid search, sehingga kombinasi bobot yang menghasilkan nilai koefisien determinasi (R^2) tertinggi pada data validasi dipilih sebagai model ensemble akhir. Secara matematis, prediksi ensemble dinyatakan sebagai:

$$\hat{y}_{ens} = w_{GB}\hat{y}_{GB} + w_{RF}\hat{y}_{RF} + w_{RBFNN}\hat{y}_{RBFNN} \quad (2)$$

dengan ketentuan:

$$w_{GB} + w_{RF} + w_{RBFNN} = 1 \quad (3)$$

di mana w_{GB} , w_{RF} , dan w_{RBFNN} merupakan bobot masing-masing model, sedangkan $\hat{y}_{GB}$, $\hat{y}_{RF}$, dan $\hat{y}_{RBFNN}$ merupakan prediksi yang dihasilkan oleh setiap algoritma. Pendekatan weighted averaging merupakan salah satu metode ensemble yang umum digunakan karena mampu mengombinasikan keunggulan beberapa model untuk meningkatkan akurasi prediksi dan mengurangi variansi model tunggal [[21],[22]].

Pengolahan Dataset

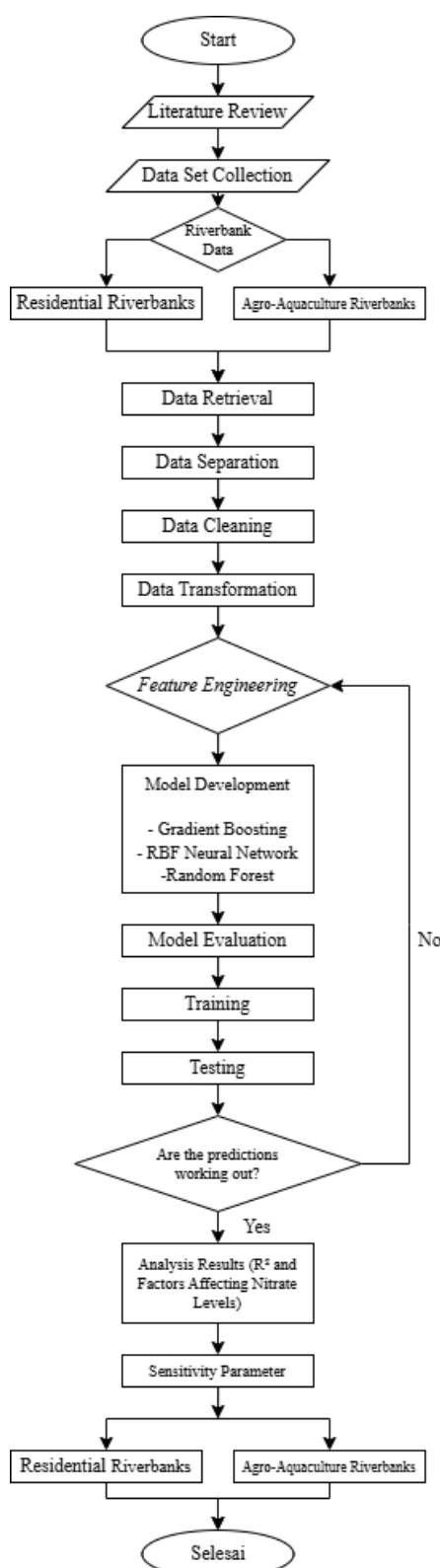
Dataset yang digunakan pada penelitian ini merupakan data kualitas air sungai yang telah dikompilasi dan diseleksi sesuai kriteria penelitian. Variabel yang digunakan terdiri atas sepuluh parameter kualitas air sebagai variabel masukan (input), yaitu suhu, pH, TDS, TSS, DO, BOD, COD, nitrit (NO_2^-), amonia (NH_3), dan DHL (Daya Hantar Listrik), sedangkan konsentrasi nitrat (NO_3^-) digunakan sebagai variabel keluaran (target). Pemilihan parameter tersebut didasarkan pada ketersediaan data yang konsisten pada seluruh observasi serta keterkaitannya dengan proses transformasi nitrogen dan karakteristik kualitas air yang memengaruhi pembentukan nitrat pada perairan.

Berdasarkan hasil kompilasi dataset, diperoleh 99 data observasi yang bersumber dari 34 artikel untuk kategori bantaran pemukiman. Selain itu, diperoleh 129 data observasi yang bersumber dari 33 artikel untuk kategori bantaran pertanian/peternakan. Jadi jumlah total observasi literatur/ artikel yang digunakan sejumlah 40 literatur. Perbedaan antara jumlah artikel dan jumlah data observasi disebabkan oleh banyaknya titik sampling pengambilan data kualitas air sungai yang dilaporkan dalam masing-masing artikel, di mana satu artikel dapat menghasilkan lebih dari satu titik data. Seluruh observasi memiliki variabel masukan dan variabel target yang lengkap sehingga dapat digunakan pada proses pemodelan. Tahap pengolahan dataset diawali dengan pembacaan data menggunakan fungsi read dataset, kemudian dilakukan pemisahan antara variabel masukan (input features) dan variabel target (target variable), yaitu konsentrasi nitrat (NO_3^-).

Selanjutnya dilakukan proses pemeriksaan kualitas data melalui konversi seluruh atribut numerik ke format numerik yang seragam. Nilai yang tidak dapat dikonversi secara langsung diperlakukan sebagai missing value, kemudian ditangani menggunakan median dari masing-masing variabel. Pendekatan ini dipilih karena median lebih robust terhadap keberadaan pencilan (outlier) dibandingkan nilai rata-rata sehingga mampu mempertahankan karakteristik distribusi data.

Untuk mengurangi pengaruh perbedaan skala antarvariabel, seluruh variabel masukan dinormalisasi menggunakan RobustScaler, yang melakukan penskalaan berdasarkan median dan Interquartile Range (IQR) sehingga lebih tahan terhadap pencilan dibandingkan metode normalisasi berbasis rata-rata dan simpangan baku. Selain itu, karena distribusi konsentrasi nitrat menunjukkan kecenderungan data yang cenderung miring ke kanan, variabel target ditransformasikan menggunakan fungsi $\log_{10}()$ untuk meningkatkan kestabilan distribusi data selama proses pembelajaran model. Tahap berikutnya adalah feature engineering, yang dilakukan dengan membentuk fitur kuadrat (square features), akar kuadrat (square-root transformation), transformasi logaritmik, serta interaksi antarvariabel (interaction features) untuk memperkaya representasi informasi yang dipelajari model.

Selanjutnya, untuk meningkatkan kemampuan generalisasi model, dilakukan augmentasi data melalui penambahan Gaussian noise dengan amplitudo 0,003 sebanyak 20 kali terhadap data pelatihan sesuai konfigurasi implementasi model. Dataset hasil pengolahan kemudian digunakan pada proses pemodelan. Untuk Hybrid RBFNN, proses pelatihan memanfaatkan validation split sebesar 90%:10%, di mana 90% data digunakan sebagai data pelatihan dan 10% sebagai data validasi untuk memantau validation loss melalui mekanisme Early Stopping dan Reduce Learning Rate on Plateau. Sementara itu, model Gradient Boosting Regressor dan Random Forest Regressor divalidasi menggunakan 5-Fold Cross Validation sehingga setiap observasi memperoleh kesempatan menjadi data validasi secara bergantian. Pendekatan ini digunakan untuk memperoleh estimasi performa model yang lebih stabil serta mengurangi potensi overfitting. Berikut ditampilkan gambar diagram alir penelitian.



Gambar 2. Diagram Alir Pengolahan Dataset RBFNN

Terdapat pula proses feature engineering dilakukan dengan membentuk fitur turunan dan interaksi antar variabel guna memperkaya informasi dalam dataset. Model kemudian dilatih

menggunakan algoritma Gradient Boosting Regressor, Random Forest Regressor, dan Radial Basis Function Neural Network (RBFNN), yang selanjutnya dikombinasikan melalui pendekatan ensemble berbasis pembobotan untuk meningkatkan akurasi prediksi. Kinerja model dievaluasi menggunakan koefisien determinasi (R^2) untuk menilai kemampuan model dalam memprediksi konsentrasi nitrat secara akurat.

Kriteria Penilaian Performa dari Pemodelan Radial Basis Function Neural Network (Hybrid RBFNN)

Kinerja model dievaluasi untuk menilai kemampuan algoritma dalam memprediksi konsentrasi nitrat berdasarkan parameter kualitas air sungai. Evaluasi dilakukan pada setiap model tunggal, yaitu Gradient Boosting Regressor (GB), Random Forest Regressor (RF), dan Hybrid Radial Basis Function Neural Network (Hybrid RBFNN), kemudian dibandingkan dengan model ensemble yang diperoleh melalui kombinasi ketiga model menggunakan pendekatan weighted averaging. Pendekatan ini bertujuan untuk mengetahui apakah penggabungan beberapa model mampu memberikan performa prediksi yang lebih baik dibandingkan penggunaan model tunggal.

Koefisien determinasi (R^2) digunakan sebagai indikator utama untuk mengukur kemampuan model dalam menjelaskan variasi data aktual. Nilai R^2 berada pada rentang $[-\infty, \infty]$, dengan nilai yang semakin mendekati satu menunjukkan bahwa model memiliki kemampuan prediksi yang semakin baik. Sebaliknya, nilai yang rendah atau bernilai negatif menunjukkan bahwa model belum mampu merepresentasikan hubungan antara variabel masukan dan keluaran secara memadai ([23] ; [24] Secara matematis, R^2 dirumuskan sebagai berikut [25]:

$$r^2 = \frac{(b_1 \sum x_1 y) + (b_2 \sum x_2 y)}{\sum y^2} \quad (4)$$

Di mana: 1) b_1 merupakan koefisien regresi untuk variabel independen pertama x_1 . Ini menunjukkan besarnya perubahan pada y jika x_1 berubah satu satuan; 2) b_2 merupakan koefisien regresi untuk variabel independen kedua x_2 yang menunjukkan besarnya pengaruh x_2 terhadap y ; 3) $\sum x_1 y$ merupakan jumlah perkalian antara selisih variabel x_1 dari rata-ratanya dengan selisih variabel y dari rata-ratanya; 4) $\sum x_2 y$ sama dengan sebelumnya, namun untuk variabel independen kedua (x_2); 5) $\sum y^2$ merupakan jumlah kuadrat penyimpangan variabel y dari rata-ratanya (Total Sum of Squares). Ini mewakili total variasi yang ada pada data asli.

Selain koefisien determinasi, penelitian ini menggunakan Mean Absolute Error (MAE) sebagai indikator kesalahan prediksi. MAE menggambarkan rata-rata selisih absolut antara nilai aktual dan nilai prediksi sehingga memberikan interpretasi yang mudah terhadap besar kesalahan model dalam satuan yang sama dengan variabel target Willmott & Matsuura [26]. Nilai MAE dihitung menggunakan persamaan:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Semakin kecil nilai MAE menunjukkan bahwa prediksi model semakin mendekati nilai observasi. Untuk mengevaluasi kemampuan generalisasi model berbasis pohon keputusan, yaitu Gradient Boosting Regressor dan Random Forest Regressor, dilakukan validasi silang menggunakan 5-Fold Cross Validation. Pada metode ini dataset dibagi menjadi lima bagian, kemudian setiap bagian digunakan secara bergantian sebagai data validasi, sedangkan empat bagian lainnya digunakan sebagai data pelatihan. Nilai performa dilaporkan sebagai rata-rata

koefisien determinasi beserta simpangan bakunya sehingga mampu memberikan gambaran yang lebih representatif terhadap stabilitas model Kohavi [27].

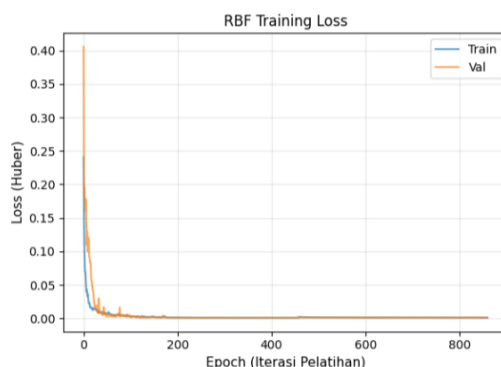
Pada penelitian ini, proses pelatihan Hybrid RBFNN menggunakan validation split sebesar 90%:10%, disertai mekanisme Early Stopping dan Reduce Learning Rate on Plateau untuk memantau nilai validation loss selama proses optimasi. Pendekatan ini dipilih untuk mencegah overfitting sekaligus mempertahankan efisiensi proses pelatihan jaringan saraf.

Selanjutnya, performa model dibandingkan berdasarkan nilai R^2 dan MAE yang diperoleh dari masing-masing model tunggal (Gradient Boosting, Random Forest, dan Hybrid RBFNN) serta model ensemble. Perbandingan tersebut digunakan untuk mengevaluasi kontribusi pendekatan ensemble learning dalam meningkatkan akurasi prediksi konsentrasi nitrat dibandingkan apabila setiap algoritma digunakan secara terpisah.

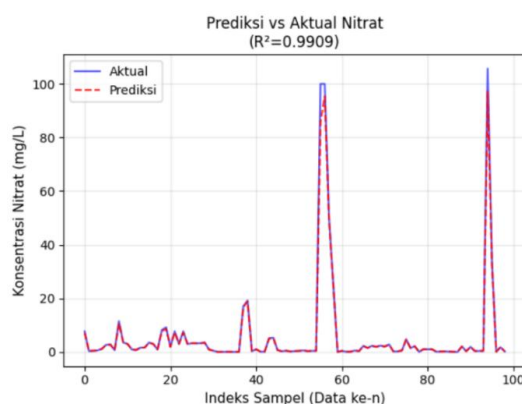
Hasil dan Pembahasan

Kinerja Model Radial Basis Function Neural Network (Hybrid RBFNN)

Berdasarkan karakteristik penggunaan lahan di sekitar aliran sungai, data kualitas air dalam penelitian ini dikelompokkan menjadi dua kategori yakni bantaran pemukiman dan bantaran agro-akuakultur, yang mencakup aktivitas pertanian, perkebunan, budidaya perairan, dan peternakan. Pengelompokan ini dilakukan untuk menguji apakah perbedaan kondisi lingkungan bantaran sungai benar-benar mempengaruhi kemampuan model untuk menangkap keterkaitan antar parameter kualitas air. Ini juga diakui dalam penelitian [28].



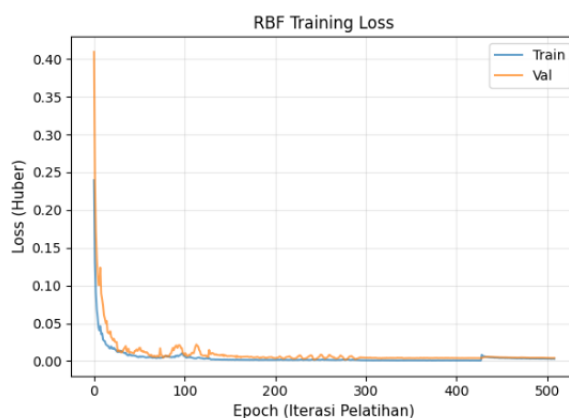
Gambar 3. Grafik *Training loss* dan *Validation loss* Bantaran Pemukiman



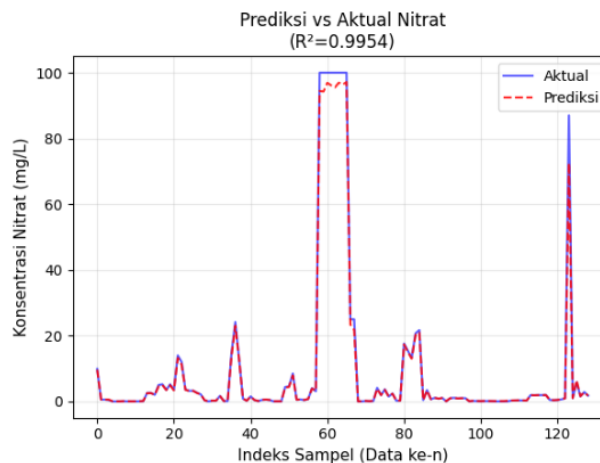
Gambar 4. Grafik Prediksi VS Aktual Nitrat Bantaran Pemukiman

Kurva training loss dan validation loss pada kedua kategori bantaran yang ditunjukkan di Gambar 3 dan Gambar 5 memperlihatkan pola yang serupa: penurunan tajam pada epoch 0–50, kemudian melandai stabil mendekati nol setelah epoch 200. Kedua kurva bergerak selaras tanpa saling menjauh, mengindikasikan bahwa model tidak mengalami overfitting dan mampu melakukan generalisasi dengan baik terhadap data baru pada kedua tipe bantaran.

Kesesuaian tersebut juga terlihat pada perbandingan nilai aktual dan prediksi yang berada di Gambar 4 dan Gambar 6, di mana model mampu mengikuti pola nitrat pada rentang nilai rendah maupun pada beberapa lonjakan ekstrem yang mendekati dan melampaui 100 mg/L, baik pada bantaran pemukiman ($R^2 = 0,9909$) maupun bantaran agro-akuakultur ($R^2 = 0,9954$). Kemampuan mengikuti lonjakan ekstrem ini menunjukkan bahwa pendekatan ensemble yang mengintegrasikan Gradient Boosting, Random Forest, dan RBFNN efektif menangkap hubungan nonlinier kompleks antara parameter kualitas air dan konsentrasi nitrat.



Gambar 5. Grafik *Training loss* dan *Validation loss* Bantaran Agro-Akuakultur



Gambar 6 Grafik Prediksi VS Aktual Nitrat Bantaran Agro-Akuakultur

Berdasarkan hasil pengukuran kelembaban dan suhu yang tercatat di Tabel 1, tingkat kelembaban di kos-kosan sekitar Kampus 1 UIN Syarif Hidayatullah Jakarta bervariasi antara 62,5% hingga 76,7%. Kosan dengan kelembaban terendah memiliki kelembaban sebesar 62,5%, sedangkan kosan dengan kelembaban tertinggi mencatatkan angka 76,7%. Rentang kelembaban ini menunjukkan bahwa sebagian besar kos-kosan berada dalam kategori kelembaban yang cukup tinggi, yang berpotensi memengaruhi kenyamanan penghuni.

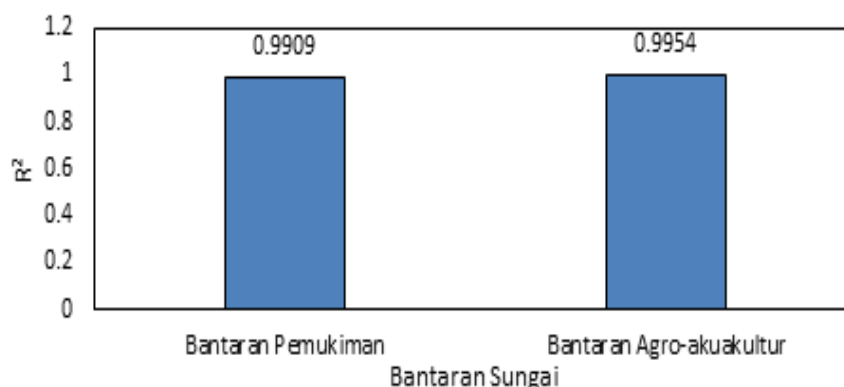
Sebagian besar kos-kosan di sekitar Kampus 1 UIN Jakarta memiliki kelembaban yang sesuai dengan standar kenyamanan termal Indonesia (terutama dalam kategori Nyaman Optimal dengan kelembaban 70% – 80%), suhu udara yang tinggi ($>25,8^{\circ}\text{C}$) menjadikan kenyamanan penghuni cenderung berada dalam kategori hampir nyaman. Hanya sedikit kos-kosan yang memenuhi standar suhu yang ideal untuk kenyamanan optimal. Oleh karena itu, meskipun kelembaban udara sebagian besar berada dalam rentang yang diinginkan, suhu udara yang tinggi dapat mengurangi kenyamanan penghuni kos.

Performa tinggi yang dicapai model tidak terlepas dari desain arsitektur Hybrid RBFNN yang digunakan. Berbeda dari RBFNN konvensional yang hanya mengandalkan lapisan radial, arsitektur pada penelitian ini menggabungkan tiga cabang pemrosesan, Dense Branch yang mempelajari hubungan nonlinier kompleks melalui fungsi aktivasi ReLU, serta Linear Branch yang mempertahankan informasi hubungan linier antarvariabel. Kombinasi ketiga representasi ini memungkinkan model menangkap karakteristik data yang beragam baik pola lokal, nonlinier, maupun linier sesuatu yang sulit dicapai oleh RBFNN tunggal. Ditambah dengan pendekatan ensemble learning berbasis weighted averaging terhadap Gradient Boosting dan Random Forest, model memperoleh keunggulan tambahan berupa reduksi variansi prediksi dan mitigasi kelemahan masing-masing algoritma tunggal [21][22].

Keberhasilan model mencapai R^2 di atas 0,99 pada kedua bantaran memiliki makna lebih dari sekadar capaian statistik. Eutrofikasi di kawasan bantaran umumnya dipicu oleh masuknya beban nitrogen berlebih ke badan air, baik dari limbah domestik pada bantaran pemukiman [29], maupun limpasan pupuk dan limbah ternak pada bantaran agro-akuakultur yang jika tidak terpantau dapat memicu ledakan alga berbahaya dan peningkatan biaya pengolahan air [30]. Akurasi model yang tinggi menunjukkan bahwa hubungan antar parameter fisika-kimia air dengan pembentukan nitrat melalui proses nitrifikasi berhasil direpresentasikan secara valid oleh model, sehingga berpotensi menggantikan pendekatan pemantauan konvensional yang bersifat reaktif dengan pendekatan yang lebih prediktif dan antisipatif.

Dari sisi pengelolaan kualitas air, keunggulan ini juga bernilai praktis karena nitrit, DHL, dan TDS jauh lebih mudah, cepat, dan murah diukur di lapangan dibandingkan analisis nitrat secara spektrofotometri di laboratorium, sehingga frekuensi pemantauan dapat ditingkatkan tanpa menambah beban biaya operasional secara signifikan. Kemampuan mendeteksi tren peningkatan nitrit dan DHL secara dini juga menjadikan kedua parameter tersebut berpotensi difungsikan sebagai indikator awal pencemaran nitrogen, sehingga upaya pengendalian pencemaran dapat dilakukan secara preventif sebelum kadar nitrat terakumulasi dan memicu eutrofikasi, alih-alih menunggu hasil analisis laboratorium yang bersifat reaktif. Nilai tambah ini penting pula ditinjau dari aspek kesehatan lingkungan, mengingat konsentrasi nitrat yang tinggi pada sumber air baku berisiko terhadap kesehatan masyarakat yang mengonsumsinya, sehingga deteksi dini melalui model ini dapat membantu meminimalkan risiko paparan tersebut.

Kinerja Radial Basis Function Neural Network (Hybrid RBFNN) Berdasarkan Karakteristik Bantaran Sungai



Gambar 7. Diagram Perbandingan Nilai Koefisien Determinasi (R^2) Model Pada Kedua Karakteristik Bantaran Sungai

Pada Gambar 7 didapat perbedaan hasil nilai R^2 pada kedua karakteristik bantaran. Hal ini dapat disebabkan dari perbedaan struktur data kualitas air dan dapat dipengaruhi oleh kombinasi model.

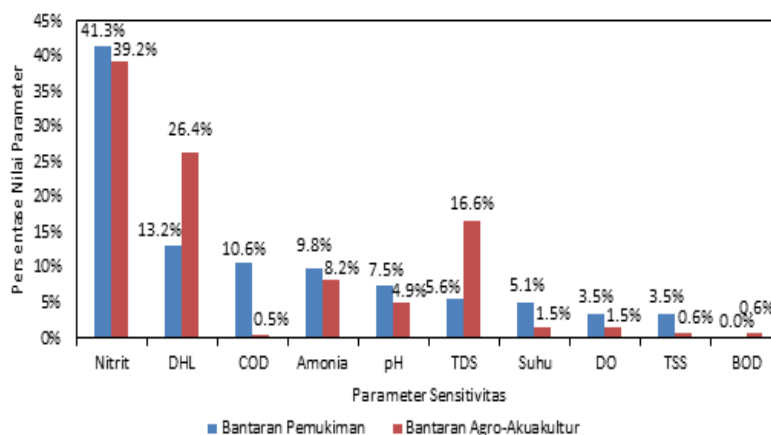
Dataset bantaran agro-akuakultur yang lebih besar (129 data) memungkinkan model mempelajari pola hubungan antar variabel secara lebih stabil dibandingkan bantaran pemukiman (99 data), sehingga model memiliki lebih banyak variasi pola untuk dipelajari secara stabil. Hal ini sejalan dengan temuan Elmotawakkil et al [31] bahwa model pembelajaran mesin memerlukan dataset yang besar dan beragam untuk menghasilkan performa yang akurat dan dapat digeneralisasi pada berbagai kondisi hidrologis.

Selain jumlah data, perbedaan sumber masukan nitrogen turut memengaruhi hubungan antar parameter. Pada bantaran pemukiman, nitrogen berasal dari limbah domestik dengan komposisi polutan yang lebih beragam [32], menghasilkan pola hubungan antarparameter yang lebih bervariasi dan sulit dipelajari secara konsisten oleh model. Sebaliknya, pada bantaran agro-akuakultur, nitrogen bersumber dari limpasan pupuk, limbah dan ternak yang menghasilkan distribusi nutrisi lebih konsisten [33], dimana hanya sekitar 39–43% nitrogen pupuk yang terserap tanaman, sisanya masuk ke badan air melalui aliran permukaan dan pencucian [34], sehingga pola pencemarannya lebih dapat diprediksi.

Model RBFNN pada penelitian ini menghasilkan nilai R^2 sebesar 0,9909 pada bantaran pemukiman dan 0,9954 pada bantaran agro-akuakultur, yang berarti model mampu menjelaskan lebih dari 99% variasi konsentrasi nitrat. Nilai ini lebih tinggi dibandingkan penelitian serupa di Indonesia, seperti pemodelan kualitas air Sungai Subayang menggunakan ANN yang menghasilkan R^2 pada kisaran 0,90–0,97 [35]. Perbedaan tersebut kemungkinan dipengaruhi oleh jumlah data pelatihan, jenis parameter, dan karakteristik lingkungan sungai [36]. Jika dibandingkan secara internasional, hasil ini juga sebanding dengan penelitian Gholamreza et al [11] yang melaporkan RBFNN mencapai R^2 sebesar 0,997 dalam memprediksi kualitas air Sungai Zayanderud di Iran. Hal ini mengindikasikan bahwa model RBFNN efektif dalam memodelkan hubungan non-linear pada sistem kualitas air sungai [37].

Analisis Sensitivitas Parameter Pada Setiap Karakteristik Bantaran

Analisis sensitivitas berfungsi untuk menilai tingkat kepekaan suatu model terhadap perubahan parameter masukan, sehingga dapat diketahui variabel input mana yang paling berkontribusi dalam menentukan nilai output [11], [38]. Berikut pada Gambar 8 menampilkan persentase hasil parameter sensitivitas kadar nitrat yang dibandingkan dengan karakteristik bantaran sungai menggunakan Hybrid RBFNN.



Gambar 8. Diagram Perbandingan Nilai Sensitivitas Dari Kedua Bantaran Sungai

Perhitungan tingkat sensitivitas masing-masing parameter dalam penelitian ini dilakukan dengan pendekatan permutation importance, yaitu dengan mengukur besarnya penurunan performa model ketika nilai suatu variabel input diacak sementara variabel lainnya tetap dipertahankan pada nilai aslinya. Semakin besar penurunan performa yang terjadi akibat pengacakan suatu variabel, semakin besar pula kontribusi variabel tersebut terhadap kemampuan model dalam memprediksi konsentrasi nitrat. Nilai persentase kontribusi pada Gambar 8 diperoleh dari rata-rata penurunan performa tersebut yang selanjutnya dinormalisasi terhadap total penurunan performa seluruh parameter, sehingga nilai yang diperoleh dapat dibandingkan secara relatif antar variabel maupun antar karakteristik bantaran sungai.

Pada Gambar 8 menunjukkan kadar nitrat (NO_3^-) sangat dipengaruhi oleh kadar nitrit, DHL (Daya Hantar Listrik), dan TDS, dengan tingkat pengaruh rata-rata sebesar 40,25% untuk nitrit. Melalui mekanisme nitrifikasi dalam siklus nitrogen di perairan, nitrit dioksidasi lebih lanjut oleh bakteri *Nitrobacter* menjadi nitrat (NO_3^-). Nitrogen menjadi kontributor pencemaran paling signifikan, terutama pupuk dan pestisida yang meresap ke air tanah dan mengalir menuju sungai pada bantaran agro-akuakultur [38]. Nilai sensitivitas terbesar kedua menunjukkan 19,80% pengaruh rata-rata pada DHL (Daya Hantar Listrik). Selanjutnya, TDS dengan rata-rata 11,10%, di mana beban nutrisi yang terbawa oleh aliran permukaan maupun limbah pada bantaran agro-akuakultur serta bantaran pemukiman masuk ke badan air sungai. Peningkatan kadar TDS dapat diakibatkan oleh ion-ion kimia yang ada di dalam aliran air sungai seperti magnesium, klorida, natrium, dan sulfat [39]. Selanjutnya diikuti ammonia sebesar 9,00%, pH sebesar 6,20%, dan COD sebesar 5,55%. Secara keseluruhan, hasil analisis sensitivitas parameter mengindikasikan bahwa konsentrasi nitrat pada sungai di kawasan tropis sangat dipengaruhi oleh dinamika kawasan bantaran sungai. Bidang agro-akuakultur serta pemukiman yang berlangsung terus-menerus di sepanjang bantaran berkontribusi langsung terhadap peningkatan beban nutrisi senyawa nitrogen yang berujung pada akumulasi nitrat di perairan sungai.

Dominannya kontribusi nitrit terhadap prediksi konsentrasi nitrat secara ilmiah dapat dijelaskan melalui mekanisme nitrifikasi dalam siklus nitrogen, yaitu proses oksidasi bertahap amonium menjadi nitrit oleh bakteri *Nitrosomonas*, yang selanjutnya dioksidasi lebih lanjut oleh bakteri *Nitrobacter* menjadi nitrat [7]. Oleh karena nitrit merupakan senyawa antara (intermediate compound) yang secara langsung berkaitan secara kimiawi dengan pembentukan nitrat, cukup beralasan apabila parameter ini memiliki kontribusi paling besar terhadap variasi nilai keluaran model dibandingkan parameter lain yang bersifat lebih tidak langsung, seperti DHL dan TDS yang lebih menggambarkan kondisi ionik atau beban zat terlarut secara umum dalam perairan.

Meskipun model Hybrid RBFNN yang dikembangkan pada penelitian ini menunjukkan performa prediksi yang tinggi, beberapa keterbatasan perlu diperhatikan dalam menginterpretasikan hasil penelitian ini.

Pertama, data yang digunakan merupakan data sekunder hasil kompilasi dari 40 publikasi yang dihasilkan melalui metode analisis laboratorium dan instrumen pengukuran berbeda-beda antarpelitian, sehingga terdapat kemungkinan variasi presisi dan akurasi pengukuran antarsumber data yang tidak sepenuhnya dapat dikendalikan dalam penelitian ini. Perbedaan metode analisis, kalibrasi alat, maupun periode pengambilan sampel antarpublikasi berpotensi menimbulkan bias sistematis yang tidak terdeteksi secara eksplisit oleh model.

Kedua, heterogenitas dataset yang berasal dari berbagai sungai tropis dengan karakteristik geografis, hidrologis, dan antropogenik yang berbeda-beda dapat memengaruhi generalisasi model. Meskipun keragaman ini secara umum bermanfaat untuk melatih model agar lebih adaptif [31], keragaman yang terlalu tinggi juga berisiko menimbulkan ketidakseimbangan representasi data antarwilayah, sehingga performa model pada wilayah dengan jumlah data yang lebih sedikit berpotensi kurang optimal dibandingkan wilayah data yang lebih dominan.

Ketiga, penelitian ini berpotensi mengandung publication bias, mengingat data yang dikumpulkan berasal dari literatur terpublikasi yang cenderung melaporkan hasil pengukuran dengan pola yang lebih jelas maupun temuan yang lebih signifikan, sedangkan data dengan hasil anomali atau tidak signifikan memiliki peluang kecil untuk dipublikasikan. Kondisi ini berpotensi memengaruhi distribusi data yang digunakan dalam pelatihan model.

Keempat, model dalam penelitian ini belum divalidasi menggunakan data lapangan independen yang dikumpulkan secara langsung di luar dataset kompilasi yang digunakan untuk pelatihan. Validasi menggunakan 5-Fold Cross Validation yang diterapkan pada penelitian ini pada dasarnya masih menggunakan populasi data yang sama, sehingga belum sepenuhnya mencerminkan kemampuan generalisasi model terhadap kondisi baru yang tidak terdapat dalam dataset pelatihan.

Kelima, penelitian ini belum melakukan validasi temporal maupun spasial secara eksplisit, seperti pengujian model pada rentang waktu pengamatan yang berbeda (misalnya musim kemarau dan musim hujan) atau pada segmen sungai tertentu yang sengaja dipisahkan sebagai data uji independen (spatial hold-out). Ketidadaan kedua validasi ini menyebabkan konsistensi model dalam menangkap variabilitas musiman maupun spasial yang lazim terjadi pada sungai tropis belum dapat dipastikan sepenuhnya. Oleh karena itu, penerapan model ini pada tahap operasional, khususnya sebagai komponen sistem peringatan dini, disarankan untuk terlebih dahulu melalui tahap validasi lapangan dan kalibrasi ulang pada lokasi target sebelum digunakan secara luas.

Kesimpulan

Penelitian ini menunjukkan bahwa model RBFNN mampu memprediksi konsentrasi nitrat (NO_3^-) pada kedua tipe bantaran sungai dengan akurasi tinggi, ditandai nilai R^2 sebesar 0,9911 pada bantaran pemukiman dan 0,9954 pada bantaran agro-akuakultur, atau setara dengan kemampuan menjelaskan lebih dari 99% variasi konsentrasi NO_3^- pada masing-masing tipe. Kontribusi ilmiah utama penelitian ini terletak pada identifikasi nitrit sebagai parameter paling berpengaruh terhadap prediksi NO_3^- (rata-rata dampak 40,25%), diikuti DHL (19,80%) dan TDS (11,10%), serta pembuktian bahwa karakteristik bantaran memengaruhi efisiensi model prediktif pada bantaran agro-akuakultur dengan pola distribusi nutrisi yang lebih teratur menghasilkan akurasi model yang lebih tinggi dibandingkan bantaran pemukiman yang memiliki sumber pencemar lebih heterogen. Temuan ini memperkuat dan sedikit melampaui capaian R^2 pada beberapa studi sejenis di Indonesia, yang mengindikasikan bahwa kombinasi volume data pelatihan dan pemilihan parameter pada penelitian ini cukup efektif untuk konteks sungai tropis.

Secara praktis, model ini menawarkan alternatif pemantauan kualitas air yang lebih efisien karena nitrit, DHL, dan TDS merupakan parameter yang relatif mudah dan murah diukur dibandingkan analisis NO_3^- laboratorium konvensional, model dapat digunakan sebagai alat bantu estimasi cepat bagi pengelola sumber daya air dalam mendeteksi indikasi pencemaran nitrat pada bantaran sungai bertipe pemukiman maupun agro-akuakultur, sekaligus mendukung pengambilan keputusan yang selaras dengan baku mutu PP No. 22 Tahun 2021.

Meskipun demikian, penelitian ini memiliki dua keterbatasan yang membuka arah pengembangan lanjutan. Pertama, jumlah literatur yang digunakan sebagai sumber data sekunder masih terbatas, sehingga pola hubungan antara karakteristik bantaran dan parameter sensitivitas yang teridentifikasi berpotensi belum sepenuhnya merepresentasikan variasi kondisi sungai tropis secara luas; oleh karena itu, penelitian mendatang disarankan memperluas basis literatur hingga lebih dari 100 jurnal guna menguji konsistensi temuan ini pada cakupan karakteristik bantaran yang lebih representatif. Kedua, model dalam penelitian ini hanya mengandalkan parameter nitrit, DHL, dan TDS sebagai prediktor utama, sehingga kontribusi parameter kualitas air lain yang berpotensi relevan seperti fosfat, klorida, sulfat, kekeruhan, dan parameter biologis belum teruji; perluasan dataset dengan parameter-parameter tersebut pada penelitian selanjutnya diperlukan agar model dapat lebih komprehensif dalam mengidentifikasi determinan konsentrasi NO_3^- di berbagai tipe bantaran sungai tropis.

Referensi

- [1] A. R. Utama, E. Yuliani, Dan T. B. Prayogo, "Analisis Kualitas Air Menggunakan Metode Indeks Pencemaran, Ccme-Wqi, Dan Qual2kw Pada Sungai Amprong, Kota Malang, Jawa Timur," *J. Teknol. Dan Rekayasa Sumber Daya Air*, Vol. 5, No. 1, Hlm. 209–219, Jan 2025, Doi: 10.21776/Ub.Jtresda.2025.005.01.021.
- [2] A. Romdani S. T., M. Eng., Ph. D., *Manajemen Sungai Dan Banjir Adaptasi Terhadap Perubahan Iklim Global*. Mega Press Nusantara, 2022.

- [3] P. Sahu, S. Londhe, Dan P. Kulkarni, "Knowledge Extraction From Trained Ann River Water Quality Parameters," *Comput. Eng. Phys. Model.*, Vol. 7, No. 4, Okt 2024, Doi: 10.22115/Cepm.2024.459617.1312
- [4] A. N. Muchamad, G. P. Prajati, T. Mulyani, Dan C. Febrion, "Optimalisasi Pemanfaatan Sumber Daya Air Dalam Wawasan Ketahanan Bangsa: Sebuah Ulasan: Sebuah Ulasan," *J. Kebangs. Ri*, Vol. 3, No. 1, Hlm. 21–31, Nov 2025, Doi: 10.31848/Jkri.V3i1.4362.
- [5] P. P. R. Mendrofa, "Analisis Fisika Kimia Kualitas Air Perairan Sungai Boyo Desa Onowaembo Kecamatan Gunungsitoli Kota Gunungsitoli," *J. Sumber Daya Akuatik*, Vol. 2, No. 1, Hlm. 1–7, Jun 2025.
- [6] T. K. Nufutomo, "Perubahan Iklim Sebagai Ancaman Ketahanan Kualitas Air Pada Daerah Aliran Sungai: Literatur Review," *J. Reka Lingkung.*, Vol. 10, No. 3, Hlm. 188–200, Nov 2022, Doi: 10.26760/Rekalingkungan.V10i3.188-200.
- [7] R. Arnanda, "Analisis Kadar Nitrat Dalam Air Sungai Dengan Menggunakan Spektrofotometer Uv-Visible," *J. Kolaboratif Sains*, Vol. 6, No. 3, Hlm. 181–184, Mar 2023, Doi: 10.56338/Jks.V6i3.3357.
- [8] Salsabila, R. Haribowo, Dan E. Yuliani, "Analisis Kualitas Air Menggunakan Metode Indeks Pencemaran, Ccme-Wqi, Dan Nsf-Wqi Di Sungai Wonokromo, Kota Surabaya, Jawa Timur," *J. Teknol. Dan Rekayasa Sumber Daya Air*, Vol. 4, No. 1, Hlm. 859–872, Jan 2024, Doi: 10.21776/Ub.Jtresda.2024.004.01.073.
- [9] M. A. Novianta Dkk., "Monitoring River Water Quality Through Predictive Modeling Using Artificial Neural Networks Backpropagation," *Aims Environ. Sci.*, Vol. 11, No. 4, Hlm. 649–664, 2024, Doi: 10.3934/Environsci.2024032.
- [11] A. Gholamreza, M.-D. Afshin, H. A. Shiva, Dan R. Nasrin, "Application Of Artificial Neural Networks To Predict Total Dissolved Solids In The River Zayanderud, Iran," *Environ. Eng. Res.*, Vol. 21, No. 4, Hlm. 333–340, Jun 2016, Doi: 10.4491/Eer.2015.096.
- [12] T. Li, J. Lu, J. Wu, Z. Zhang, Dan L. Chen, "Predicting Aquaculture Water Quality Using Machine Learning Approaches," *Water*, Vol. 14, Hlm. 2836, Sep 2022, Doi: 10.3390/W14182836.
- [13] Dodig, E. Ricci, G. Kvascev, Dan M. Stojkovic, "A Novel Machine Learning-Based Framework For The Water Quality Parameters Prediction Using Hybrid Long Short-Term Memory And Locally Weighted Scatterplot Smoothing Methods," *Journal Of Hydroinformatics*, Vol. 26, No. 5, Hlm. 1059–1079, Mei 2024, Doi: 10.2166/Hydro.2024.273.
- [14] S. S. Haykin, *Neural Networks And Learning Machines*, 3. Ed. New York Munich: Prentice-Hall, 2009.

-
- [15] C. M. Bishop, Pattern Recognition And Machine Learning. Dalam Information Science And Statistics. New York: Springer, 2006.
- [16] J. Park Dan I. W. Sandberg, "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation*, Vol. 3, No. 2, Hlm. 246–257, Jun 1991, Doi: 10.1162/Neco.1991.3.2.246.
- [17] J. Moody Dan C. J. Darken, "Fast Learning In Networks Of Locally-Tuned Processing Units," *Neural Computation*, Vol. 1, No. 2, Hlm. 281–294, Jun 1989, Doi: 10.1162/Neco.1989.1.2.281.
- [18] A. D. Jagtap Dan G. E. Karniadakis, "How Important Are Activation Functions In Regression And Classification? A Survey, Performance Comparison, And Future Directions," *J Mach Learn Model Comput*, Vol. 4, No. 1, Hlm. 21–75, 2023, Doi: 10.1615/Jmachlearnmodelcomput.2023047367.
- [19] D. P. Kingma Dan J. Ba, "Adam: A Method For Stochastic Optimization," 30 Januari 2017, Arxiv: Arxiv:1412.6980. Doi: 10.48550/Arxiv.1412.6980.
- [20] P. J. Huber, "Robust Estimation Of A Location Parameter," *Ann. Math. Statist.*, Vol. 35, No. 1, Hlm. 73–101, Mar 1964, Doi: 10.1214/Aoms/1177703732.
- [21] T. G. Dietterich, "Ensemble Methods In Machine Learning," Dalam Multiple Classifier Systems, Vol. 1857, Dalam Lecture Notes In Computer Science, Vol. 1857. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, Hlm. 1–15. Doi: 10.1007/3-540-45014-9_1.
- [22] K. Shaga Devan, H. A. Kestler, C. Read, Dan P. Walther, "Weighted Average Ensemble-Based Semantic Segmentation In Biological Electron Microscopy Images," *Histochem Cell Biol*, Vol. 158, No. 5, Hlm. 447–462, Nov 2022, Doi: 10.1007/S00418-022-02148-3.
- [23] F. Sohil, M. U. Sohali, Dan J. Shabbir, "An Introduction To Statistical Learning With Applications In R: By Gareth James, Daniela Witten, Trevor Hastie, And Robert Tibshirani, New York, Springer Science And Business Media, 2013, \$41.98, Eisbn: 978-1-4614-7137-7," *Statistical Theory And Related Fields*, Vol. 6, No. 1, Hlm. 87–87, Jan 2022, Doi: 10.1080/24754269.2021.1980261.
- [24] Hastie T, Tibshirani R, Friedman J. The Elements Of Statistical Learning: Data Mining, Inference, And Prediction. 2nd Ed. New York: Springer; 2009. <https://doi.org/10.1007/978-0-387-84858-7>
- [25] S. D. Sinambela, S. Ariswoyo, Dan H. R. Sitepu, "Menentukan Koefisien Determinasi Antara Estimasi M Dengan Type Welsch Dengan Least Trimmed Square Dalam Data Yang Mempunyai Pencilan," *Saintia Matematika*, Vol. 2, No. 3, Hlm. 225–235, 2014.

- [26] C. Willmott Dan K. Matsuura, "Advantages Of The Mean Absolute Error (Mae) Over The Root Mean Square Error (Rmse) In Assessing Average Model Performance," *Clim. Res.*, Vol. 30, Hlm. 79–82, 2005, Doi: 10.3354/Cr030079.
- [27] Kohavi R. A Study Of Cross-Validation And Bootstrap For Accuracy Estimation And Model Selection. In: *Proceedings Of The Fourteenth International Joint Conference On Artificial Intelligence (Ijcai)*. Montreal, Canada; 1995. P. 1137–1143.
- [28] Sa'adilla Mh, Saptomo Sk, Kurniawan A. Artificial Neural Network Modeling To Detect Fe Concentration In Water. *Iop Conference Series: Earth And Environmental Science*. 2025;1515(1):012012. <https://doi.org/10.1088/1755-1315/1515/1/012012>
- [29] K. A. Locke, "Impacts Of Land Use/Land Cover On Water Quality: A Contemporary Review For Researchers And Policymakers," *Water Qual. Res. J.*, Hlm. Wqrj2024002, Apr 2024, Doi: 10.2166/Wqrj.2024.002.
- [30] Sitote Ym, Gebremedhine Mg. Comprehensive Review Of Eutrophication In Freshwater Ecosystems: Causes, Effects, Assessment, And Management Strategies. *Preprints*. 2024. Available From: <https://www.preprints.org/manuscript/202409.0704>
- [31] A. Elmotawakkil, N. Enneya, S. K. Bhagat, M. M. Ouda, Dan V. Kumar, "Advanced Machine Learning Models For Robust Prediction Of Water Quality Index And Classification," *J. Hydroinformatics*, Vol. 27, No. 2, Hlm. 299–319, Feb 2025, Doi: 10.2166/Hydro.2025.290.
- [32] H. Filali, N. Barsan, D. Souguir, V. Nedeff, C. Tomozei, Dan M. Hachicha, "Greywater As An Alternative Solution For A Sustainable Management Of Water Resources—A Review," *Sustainability*, Vol. 14, No. 2, Hlm. 665, Jan 2022, Doi: 10.3390/Su14020665.
- [33] L. Yalan, P. Bo, T. Li, M. Chang, Y. Guo, Y. Liu, And X. Nie, "Agricultural Activities And Hydrological Processes Drive Nitrogen Pollution And Transport In Polder Waters: Evidence From Hydrochemical And Isotopic Analysis," *Water*, Vol. 17, No. 17, Art. No. 2601, Sep. 2025, Doi: 10.3390/W17172601
- [34] H. Xu, X. Tan, J. Liang, Y. Cui, Dan Q. Gao, "Impact Of Agricultural Non-Point Source Pollution On River Water Quality: Evidence From China," *Front. Ecol. Evol.*, Vol. 10, Apr 2022, Doi: 10.3389/Fevo.2022.858822.
- [35] E. A. Damanhuri, Y. I. Siregar, Dan E. Elfizar, "Penerapan Model Berbasis Artificial Neural Network Untuk Memprediksi Kualitas Air Di Sungai Subayang Kabupaten Kampar," *J. Ilmu Lingkung.*, Vol. 14, No. 1, Hlm. 18–28, Mar 2020, Doi: 10.31258/Jil.14.1.P.18-28.
- [36] A. Elsayed, J. Levison, A. Binns, M. Larocque, Dan P. Goel, "Regression-Based Machine Learning Models For Nitrate And Chloride Prediction In Surface Water In A Small Agricultural Sand Plain Sub-Watershed In Southwestern Ontario,"

- Canada,” *Front. Environ. Sci.*, Vol. 13, Hlm. 1543852, Mar 2025, Doi: 10.3389/Fenvs.2025.1543852.
- [37] Md. A. A. M. Hridoy Dkk., “Advanced Machine Learning Models For Accurate Water Quality Classification And Wqi Prediction: Implications For Aquatic Disease Risk Management,” *Sci. Total Environ.*, Vol. 1008, Hlm. 180965, Des 2025, Doi: 10.1016/J.Scitotenv.2025.180965.
- [38] T. D. Banda And M. Kumarasamy, "Artificial Neural Network (Ann)-Based Water Quality Index (Wqi) For Assessing Spatiotemporal Trends In Surface Water Quality A Case Study Of South African River Basins," *Water*, Vol. 16, No. 11, P. 1485, May 2024, Doi: 10.3390/W16111485.
- [39] S. Nemati, L. Naghipour, And M. H. F. Fard, "Artificial Neural Network Modeling Of Total Dissolved Solids In The Simineh River, Iran," *International Journal Of Environmental Research*, Vol. 8, No. 4, Pp. 1203–1212, 2014.