Machine Learning-Based Prediction of Received Signals in Molecular Communication Systems

Gaurav Anand, Adarsh Kumar, Dr. Lokendra Chouhan,Senior Member,IEEE

***Abstract*— This study focuses on predicting signal reception in molecular communication systems utilizing cylindrical transmitters and receivers, a design with notable advantages over spherical models in real-world scenarios. Cylindrical structures are more representative of biological and nanotechnological applications, such as nerve cells, blood vessels, and microfluidic devices. By combining simulated data with machine learning algorithms—including Linear Regression, Random Forest Regression, Support Vector Regression (SVR), and XGBoost Regression—we validate and refine the theoretical formula for cylindrical geometries. Our results demonstrate the potential of cylindrical models to improve the efficiency and applicability of molecular communication systems, paving the way for advancements in this emerging field.**

***Index Terms*— Molecular communication, machine Learning, curve fitting, efficiency and error analysis.**

I. Introduction

Molecular communication (MC) introduces an intriguing concept: the utilization of chemical molecules as a means to convey information. Picture minuscule devices, barely visible to the naked eye, operating on scales measured in microns or smaller, communicating through this innovative medium. It's not merely a theoretical notion; nature has long employed molecular communication, facilitating interactions among organisms and cells [2]. Gaining access to the futuristic potential of nanotechnology requires the ability to communicate at the nanoscale. The basic components for nanomachines have been accessible for a while now. One characteristic unites the revolutionary uses of nanotechnology: instead of involving lone devices operating separately, they involve swarms of devices cooperating (a network of nanomachines).

One of its innovative ideas is to promote collaboration among small entities. Effective communication at the nano- and micro-scale is crucial for unlocking the transformative potential of nanotechnology [2]. The received signal in an MC via diffusion (MCvD) system is made up of molecules that are emitted by a transmitter and travel through the surrounding environment to reach the receiver node. Creating reliable models to reflect the received signal under various environmental and conditional circumstances is one of the primary problems in molecular communication (MC).

The estimated cumulative number of molecules received when the transmitter is a point source and the receiver is an absorbing spherical node in a three-dimensional medium is reported by the authors in [4]. The estimated cumulative number of received molecules when the receptor effect was added to the system described in [4] was determined by the authors in [1] using an analytical modeling of the received signal.

In order to model the diffusion channel for a more realistic topology, we suggest a machine learning-based method in this study [3].

II. Related Work & Motivation

Machine learning-based prediction of received signals in molecular communication systems holds significant importance due to several key reasons:

Enhanced Accuracy and Reliability: Traditional methods of predicting received signals in molecular communication rely on simplistic models or approximations that may not capture the complex dynamics of molecular interactions accurately. Machine learning algorithms can learn from large datasets and adaptively adjust models, leading to more accurate predictions of received signals under varying conditions.

Handling Complex Relationships: Molecular communication systems often involve intricate relationships between transmitter characteristics, channel conditions, environmental factors, and receiver responses. Machine learning models excel in capturing these nonlinear and complex relationships, which are challenging to model using conventional analytical techniques.

Adaptability to Dynamic Environments: Molecular communication environments can be highly dynamic, with variables such as molecular concentration, diffusion rates, and interference levels changing over time. Machine learning models can continuously learn and update their predictions based on real-time data, making them adaptable to these dynamic conditions.

Optimization of System Performance: Predicting received signals accurately is crucial for optimizing the performance of molecular communication systems. Machine learning algorithms can optimize parameters such as transmission rates, signal modulation techniques, and receiver configurations to maximize communication reliability and efficiency based on predicted signal characteristics.

Facilitating System Design and Deployment: By providing accurate predictions of received signals, machine learning enables better design and deployment of molecular communication systems. Engineers can simulate different scenarios, evaluate performance metrics, and fine-tune system parameters before actual implementation, thus reducing trial-and-error in system development.

*This work was not supported by any organization

[1]H. Kwakernaak is with Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, 7500 AE Enschede, The Netherlands h.kwakernaak at papercept.net

[2]P. Misra is with the Department of Electrical Engineering, Wright State University, Dayton, OH 45435, USA p.misra at ieee.org

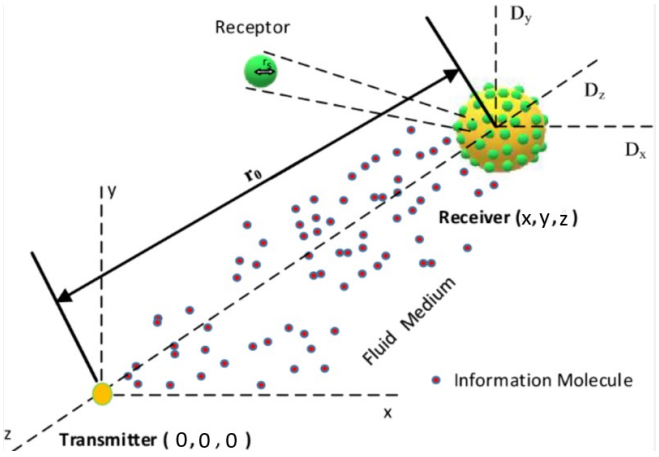

Fig. 1. Molecular communication via point transmitter and spherical receiver

Potential for Autonomous Systems: As machine learning models become more sophisticated, there is potential for autonomous molecular communication systems where predictive models enable real-time decision-making and adaptation without human intervention. This can lead to self-optimizing networks capable of adjusting to changing environmental conditions and operational requirements.

Support for Multi-Scale Modeling: Molecular communication spans multiple scales, from molecular interactions to macroscopic system behavior. Machine learning can integrate data across these scales, allowing for holistic modeling and prediction that encompasses the entire communication process.

In essence, machine learning-based prediction of received signals in molecular communication systems not only improves the accuracy and reliability of signal predictions but also opens up opportunities for optimizing system performance, adapting to dynamic environments, and advancing towards autonomous communication systems in the future.

III. Contribution

In this study, we employ a simulation-based training approach, generating a synthetic dataset that models molecular communication systems with cylindrical geometries. These cylindrical transmitter and receiver models are designed to better align with real-world environments, such as biological systems and nanotechnological applications. By applying machine learning algorithms—including Linear Regression, Random Forest Regression, Support Vector Regression (SVR), and XGBoost Regression—we validate the theoretical formulas for signal reception. Our findings highlight that cylindrical geometries provide more accurate predictions of molecule reception at the receiver, emphasizing their superior alignment with practical scenarios and their potential to enhance the reliability of molecular communication systems.

IV. Proposed Methodology

To model the diffusion of molecules in cylindrical coordinates, we start by defining the diffusion equation, set the boundary and initial conditions, and use Green's function for the solution. The integral solution is then simplified, resulting in the final expression for the fraction of molecules hitting the receiver by time t.

A. Diffusion Equation in Cylindrical Coordinates

The diffusion equation in cylindrical coordinates, assuming azimuthal symmetry, is given by:

$$\frac{\partial C(r,z,t)}{\partial t} = D\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial C}{\partial r}\right) + \frac{\partial^2 C}{\partial z^2}\right)$$

Here:

- $C(r,z,t)$: concentration of molecules at point (r,z) and time t,
- D: diffusion coefficient.

B. Boundary and Initial Conditions

The initial and boundary conditions for the system are:

1) **Transmitter Initial Condition:**

$$C(r,z,0) = C_0 \quad for 0 \le r \le r_{Tx},\ 0 \le z \le h_{Tx}$$
$$C(r,z,0) = 0 \quad for r > r_{Tx}.$$

2) **Receiver Boundary Condition:**

$$C(r_{Rx},z,t) = 0 \quad for 0 \le z \le h_{Rx}.$$

C. Green's Function for Diffusion

The Green's function $G(r,z,t|r_0,z_0)$ for diffusion in cylindrical coordinates is:

$$G(r,z,t|r_0,z_0) = \frac{1}{4\pi Dt}\exp\left(-\frac{(r-r_0)^2}{4Dt}\right)\exp\left(-\frac{(z-z_0)^2}{4Dt}\right)$$

Here:

- (r_0,z_0): initial position of molecules,
- t: time,
- $\exp(\cdot)$: exponential function.

D. First-Hitting Probability

The fraction of molecules that hit the receiver by time t is:

$$F_{hit}(t) = \int_0^{h_{Rx}}\int_0^{r_{Rx}} G(r,z,t|r_0,z_0)\, r\, dr\, dz$$

Substituting the Green's function, we get:

$$F_{hit}(t) = \frac{1}{4\pi Dt}\int_0^{h_{Rx}}\int_0^{r_{Rx}} \exp\left(-\frac{(r-r_0)^2}{4Dt}\right) r\, dr \exp\left(-\frac{(z-z_0)^2}{4Dt}\right)$$

E. Solving the Integrals

1) **Radial Integral:**

$$\int_0^{r_{Rx}} \exp\left(-\frac{(r-r_0)^2}{4Dt}\right) r\, dr = C{\cdot}erfc\left(\frac{r_{Rx}-r_0}{2\sqrt{Dt}}\right)$$

where:

- $erfc(\cdot)$: complementary error function,
- C: constant.

2) **Axial Integral:**

$$\int_0^{h_{Rx}} \exp\left(-\frac{(z-z_0)^2}{4Dt}\right) dz = \sqrt{4Dt}{\cdot}erf\left(\frac{h_{Rx}-z_0}{2\sqrt{Dt}}\right)$$

where:

- $erf(\cdot)$: error function.

F. Final Expression

Combining the radial and axial integrals, the final expression for $F_{hit}(t)$ is:

$$F_{hit}(t) = \frac{1}{4\pi Dt} \cdot C \cdot erfc\left(\frac{r_{Rx} - r_0}{2\sqrt{Dt}}\right) \cdot \sqrt{4Dt} \cdot erf\left(\frac{h_{Rx} - z_0}{2\sqrt{Dt}}\right)$$

Here:

- r_{Rx}: receiver radius,
- h_{Rx}: receiver height,
- r_0, z_0: initial positions of the molecules.

V. MODEL TRAINING

To train and test the models, a synthetic dataset was generated, incorporating a comprehensive set of parameters essential for simulating a molecular communication environment. This approach ensures that the models are trained and evaluated under conditions that closely mimic real-world molecular communication scenarios, capturing the complexity and variability of such systems

A. Dataset Generation

As shown in TABLE 1, using these parameters, we generated 10,000 random samples to create a robust dataset for training and testing our models. In addition to the primary

Parameter	Defined Range
Distance	10 µm to 100 µm
Radius	0.1 µm to 1 µm
Diffusion Coefficient	1e-12 m²/s to 100e-12 m²/s
Time	1 s to 10 s
$Deltat$	0.001 s to 0.01 s
N (Number of Molecules)	500 to 10,000

TABLE I
PARAMETERS AND THEIR DEFINED RANGES FOR DATASET GENERATION

parameters, we derived an additional column to enhance the dataset.

This approach ensures that the dataset is comprehensive and reflects the various conditions that can affect molecular communication systems. The generated data provides a solid foundation for training and testing machine learning models, enabling accurate predictions and insights into the behavior of molecular signals.

B. Data Splitting

Data splitting is a crucial step in the machine learning workflow. It ensures that our models are trained and evaluated effectively, thereby preventing overfitting and ensuring generalizability to new, unseen data. For this project, the generated dataset was split into training and testing sets, maintaining a balance between learning from the data and evaluating the model's performance.

- **Training Set(80%):** The training set is used to train the machine learning models. By exposing the models to a large and varied portion of the data, we aim to capture the underlying patterns and relationships between the input parameters and the target variable.
- **Testing Set(20%):** The testing set is reserved for evaluating the performance of the trained models. By using a separate subset of data that the models have not seen during training, we can assess how well the models generalize to new data.

To ensure that the data splitting process was performed correctly and randomly, we employed the **train_test_split** function from the scikit-learn library. This function shuffles the data before splitting, which helps in achieving a representative distribution of samples in both the training and testing sets.

C. Applying algorithms to model

To enhance the predictive accuracy of the cumulative number of received molecules over time, we employed four distinct machine learning algorithms on the simulated data. Each algorithm was selected based on its unique strengths and relevance to the task at hand. By utilizing a diverse set of algorithms, we aimed to comprehensively explore the dynamics of molecular communication systems and identify the most effective approach for accurate prediction.

- **Linear Regression:** Linear regression is a fundamental supervised learning algorithm used to model the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the input features and the target variable. LinearRegression() is imported from the‘**sklearn.linear_model**‘ library.
- **Random Forest Regression:** Random forest regression is an ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction of the individual trees. It is effective for capturing complex nonlinear relationships between features and target variables. RandomForestRegressor() is imported from the‘**sklearnėnsemble**‘ library.
- **Support Vector Regression:** Support vector regression (SVR) is a variant of support vector machines (SVMs) used for regression tasks. It works by finding the hyperplane that best fits the data while minimizing the error within a specified tolerance level. SVR() is imported from the ‘**sklearnṡvm**‘ library.
- **XGBoost Regression:** XGBoost (Extreme Gradient Boosting) is an efficient and scalable implementation of gradient boosting machines. It builds multiple decision trees sequentially, with each tree correcting the errors of its predecessors. XGBoost is known for its high performance and ability to handle complex datasets. XGBRegressor() is imported from the ‘**xgboost**‘ library.

All of these above mentioned algorithms use two function for training and prediction of the model.

- **fit(X, y):** For training the respective models.
- **predict(X):** Makes predictions using the trained model.

These functions, along with their respective libraries, are utilized for training machine learning models and making predictions in the molecular communication system. They offer a diverse set of techniques for capturing underlying

patterns and relationships in the data, enabling accurate and reliable predictions of the cumulative number of received molecules.

VI. RESULTS AND ANALYSIS

A. *Received Signal Analysis*

In the following Fig. 2, the graph illustrates the relationship between the number of received molecules and time for a distance of 2 µm. The analysis compares the performance of molecular communication systems modeled using three transmitter geometries: **point**, **spherical**, and **cylindrical**.

The cylindrical transmitter demonstrates superior performance compared to the point and spherical geometries. This improvement is further validated through machine learning models, which identify consistent patterns in the data and verify the accuracy of the proposed approach.

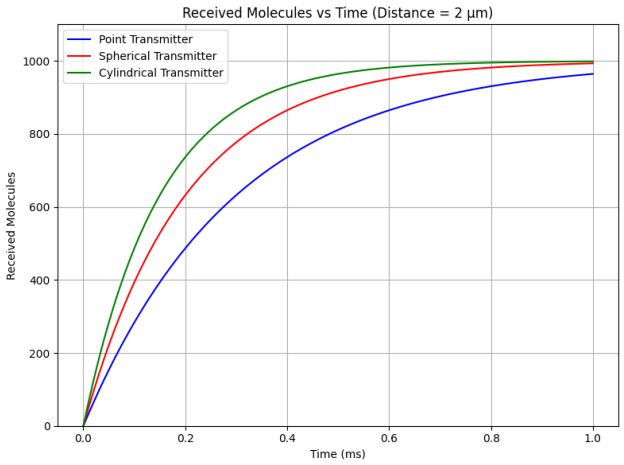

Fig. 2. Received Molecules vs Time (Distance = 2 µm)

The graph shows the following key observations:

- **Point Transmitter (Blue Curve):** The point transmitter exhibits the slowest rise in the number of received molecules. This is due to the isotropic diffusion nature, where molecules disperse evenly in all directions, resulting in fewer molecules reaching the receiver over time.
- **Spherical Transmitter (Red Curve):** The spherical transmitter shows an earlier peak compared to the point transmitter. This is attributed to the larger surface area, which facilitates faster initial molecule dispersion. However, the peak flattens earlier due to boundary effects limiting the molecule concentration near the receiver.
- **Cylindrical Transmitter (Green Curve):** The cylindrical transmitter demonstrates the steepest rise and highest peak among the three geometries. This improvement is a result of its elongated shape, which aligns the molecular diffusion direction more effectively toward the receiver. The concentration of molecules at the receiver remains higher over time, indicating improved efficiency.

Additionally, the machine learning models trained on these data sets validate the superiority of the cylindrical transmitter by accurately predicting its performance with reduced error. The models consistently highlight the faster rise and higher peak of the cylindrical transmitter's curve, reinforcing the effectiveness of the proposed design.

B. *Error Analysis*

- **Mean Squared error:** From the Fig. 4, we can easily depict that XGBoost Regression appears to have the lowest mean squared error and Random Forest regression is comparatively low where as SVR appears to have the highest mean square error.

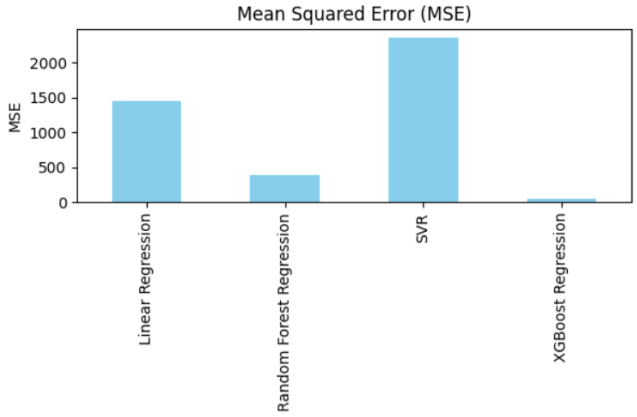

Fig. 3. Mean Squared Error(MSE)

- **Mean Absolute Error:** From Fig. 5, we can analyse that XGBoost Regression appears to have the lowest mean absolute error where as linear regression and Random forest also have relatively low MAE leaving SVR as highest.

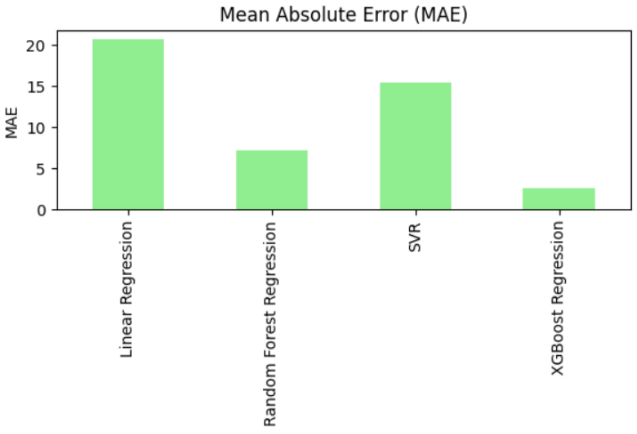

Fig. 4. Mean Absolute Error(MAE)

- **Root Mean Squared Error:** From Fig. 6, we comparatively have XGBoost Regression as the lowest root mean square error where as SVR as the highest.Linear regression is also high.
- **R-Squared:** Here SVR have the lowest R-squared followed by Linear Regression which is comparatively low

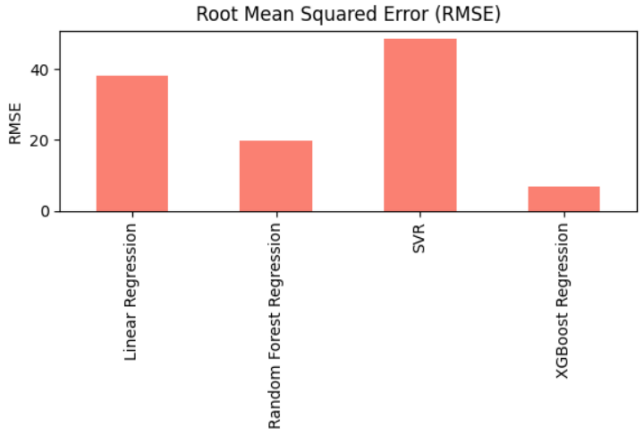

Fig. 5. Root Mean Squared Error(RMSE)

leaving Random Forest and XGBoost as the highest R-squared.

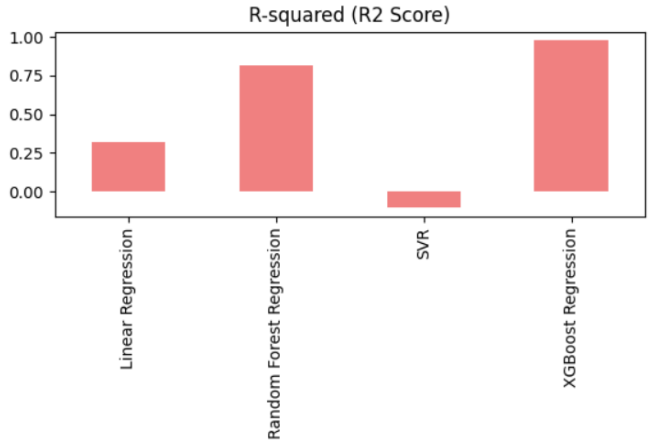

Fig. 6. R-Squared(R2 Score)

In our pursuit of accurately modeling the received signal in molecular communication, we conducted a meticulous error analysis utilizing four fundamental metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2 Score). By employing these error metrics, we gained comprehensive insights into the performance of our predictive models for molecular communication. This analysis facilitated the selection of the most effective model for accurately modeling the received signal in the communication system.

VII. CONCLUSIONS

In this study, the received signal in molecular communication systems was analyzed using point, spherical, and cylindrical transmitters. Our results indicate that the cylindrical transmitter outperformed both the point and spherical transmitters in terms of received signal strength and temporal efficiency. This demonstrates the efficacy of cylindrical transmitter modeling in enhancing molecular communication performance.

The performance of the cylindrical model was further validated by applying various machine learning algorithms, including Linear Regression, Random Forest, SVR, and XGBoost. The machine learning models effectively captured the underlying patterns of molecular diffusion, providing reliable predictions that aligned with the experimental and theoretical results. This dual-layered approach of analytical modeling and machine learning verification underscores the robustness of the cylindrical system in molecular communication.

Additionally, the error analysis revealed consistent improvements in prediction accuracy for the cylindrical transmitter, confirming its viability for real-world applications. The findings open new possibilities for utilizing cylindrical transmitter-receiver models in advanced applications, such as nanorods for targeted drug delivery and other biomedical systems.

Future research can build upon this work by exploring dynamic and real-time adaptive algorithms, incorporating environmental factors such as varying diffusion coefficients, and extending the analysis to multi-receiver systems. These enhancements could further optimize molecular communication networks and pave the way for revolutionary applications in nanotechnology, biomedical devices, and microfluidic systems.

REFERENCES

[1] Ali Akkaya, H Birkan Yilmaz, Chan-Byoung Chae, and Tuna Tugcu. Effect of receptor density and size on signal reception in molecular communication via diffusion with an absorbing receiver. *IEEE Communications Letters*, 19(1):155–158, 2014.

[2] Weisi Guo, Taufiq Asyhari, Nariman Farsad, H Birkan Yilmaz, Bin Li, Andrew Eckford, and Chan-Byoung Chae. Molecular communications: Channel model and physical layer techniques. *IEEE Wireless Communications*, 23(2):120–127, 2016.

[3] Amalendu Patnaik, Dimitrios E Anagnostou, Rabindra K Mishra, JC Lyke, et al. Applications of neural networks in wireless communications. *IEEE Antennas and Propagation Magazine*, 46(4):130–137, 2004.

[4] H Birkan Yilmaz, Akif Cem Heren, Tuna Tugcu, and Chan-Byoung Chae. Three-dimensional channel characteristics for molecular communications with an absorbing receiver. *IEEE Communications Letters*, 18(3):929–932, 2014.

REFERENCES