Predictive Modeling of Healthcare Traffic Using Machine Learning: A Comparative Study

Shadman Mahmood Khan Pathan, Sakan Binte Imran, M M Shabab Iqbal, M M Shabab Iqbal, Muhammad Enayetur Rahman, Md Nurul Absar Siddiky, Muhammad Rezaur Rahman, Md Rafid Hasan, Nondon Lal Dey, MD Sobuj Hossain

DOI: 10.29322/IJSRP.14.09.2024.p15305

Paper Received Date: 13th July 2024
Paper Acceptance Date: 26th August 2024
Paper Publication Date: 6th September 2024

Abstract- Effective healthcare traffic management is critical for ensuring prompt medical services, particularly in emergencies where delays can have life-threatening consequences. This study conducts a comparative analysis of three popular machine learning models—Linear Regression, Decision Trees, and Random Forests—for predicting healthcare-related traffic volumes. Utilizing a comprehensive dataset from a metropolitan interstate traffic system, the models were evaluated based on key performance metrics, including Mean Squared Error (MSE), R^2 Score, and execution time. The findings demonstrate that the Random Forest model outperforms the others, offering superior predictive accuracy and efficiency. These insights are valuable for optimizing traffic management in healthcare, ultimately contributing to improved patient outcomes.

Index Terms- Healthcare traffic management, machine learning, Random Forest, traffic prediction, emergency medical services, predictive modeling, urban traffic systems.

I. Introduction

Effective traffic management is crucial in healthcare, where timely access to medical facilities directly impacts patient outcomes. Urban traffic systems have grown increasingly complex, posing significant challenges to traditional traffic forecasting methods. These conventional approaches often struggle to manage the nonlinear, dynamic nature of traffic flow, which is influenced by a multitude of factors including weather conditions, time of day, and emergency incidents. In the context of healthcare, where every second counts, the limitations of traditional traffic forecasting methods become particularly pronounced. The unpredictability of traffic can lead to delays in emergency medical services (EMS), potentially compromising patient care and outcomes.

To address these challenges, the application of machine learning (ML) models in traffic forecasting has gained significant traction. ML models excel at handling large datasets and capturing complex, nonlinear patterns that are characteristic of traffic systems. Zhang and Liu (2009) highlighted the potential of machine learning methods in traffic forecasting, demonstrating that these approaches can effectively learn from historical data and improve the accuracy of traffic predictions. Their work laid the foundation for the application of advanced ML techniques in various domains, including transportation and healthcare.

One of the pioneering contributions to predictive modeling is Breiman's (1984) development of Classification and Regression Trees (CART). This method provided a robust framework for making predictions by recursively partitioning data into subsets, a technique that has been widely adopted across various fields. Breiman's subsequent work on Random Forests (2001) marked a significant advancement in ensemble learning methods. Random Forests, an ensemble of decision trees, are particularly effective in improving predictive accuracy by reducing the variance associated with single decision trees. This method has become a cornerstone in predictive modeling, with widespread applications in traffic forecasting, as demonstrated by Wu et al. (2018). Their study on freeway traffic flow prediction underscored the robustness of ensemble learning methods in handling the variability and uncertainty inherent in real-world traffic data.

In healthcare, accurate traffic forecasting has the potential to optimize the allocation of emergency medical services, thereby reducing response times and improving patient outcomes. For instance, advanced ML models can be integrated into EMS dispatch systems to predict traffic conditions in real-time, enabling ambulances to navigate through congested areas more efficiently. This integration could also facilitate better hospital logistics, ensuring that resources are allocated based on anticipated traffic patterns. However, despite the evident advantages, the application of machine learning models specifically for healthcare traffic management remains underexplored.

Several studies have focused on the broader context of traffic flow prediction using machine learning. Lv et al. (2015) explored the use of deep learning techniques for traffic flow prediction, demonstrating their effectiveness in handling large-scale, high-dimensional data. Their work emphasized the importance of leveraging big data to improve the accuracy of traffic predictions, an approach that is highly relevant to the healthcare sector, where timely decision-making is crucial. Moreover, Pathan and colleagues have made significant contributions to the application of ML and deep learning in various domains, including healthcare and agricultural systems. For instance, Pathan and Ali (2019) implemented Faster R-CNN for paddy plant disease recognition, showcasing the potential of ML in critical areas requiring real-time decision-making. Similarly, Pathan et al. (2020) developed a wireless head gesture-controlled robotic wheelchair, highlighting the role of ML in enhancing the quality of life for physically disabled persons through innovative, responsive systems.

Building on these studies, this paper aims to bridge the gap in healthcare traffic management by conducting a comparative analysis of three widely used machine learning models—Linear Regression, Decision Trees, and Random Forests—in predicting healthcare-related traffic volumes. The choice of models is informed by their varying complexity and computational efficiency. Linear Regression serves as a baseline, offering simplicity and interpretability, while Decision Trees and Random Forests provide insights into the benefits of non-linear modeling and ensemble learning in capturing the intricacies of traffic data.

The novelty of this study lies in its focus on the application of these models in a healthcare context, using a real-world dataset to evaluate their performance. By leveraging a dataset that reflects actual traffic conditions, the study provides practical insights into how different models perform under scenarios pertinent to healthcare. This work not only contributes to the existing body of knowledge but also offers actionable insights for healthcare providers and city planners seeking to improve the efficiency of medical services in urban settings.

The contributions of this study are twofold. First, it provides a comprehensive comparison of the predictive capabilities of Linear Regression, Decision Trees, and Random Forests in a healthcare traffic scenario. This comparison is crucial for identifying the most effective model for specific applications within the healthcare sector. Second, the study offers practical insights into the selection of appropriate models for traffic management in medical emergencies, emphasizing the importance of balancing accuracy with computational efficiency in real-time applications. By addressing these aspects, the study aims to enhance the understanding of how machine learning can be effectively applied to optimize healthcare traffic management, ultimately contributing to better patient outcomes.

II. Materials and Methods

1.1. Dataset

The primary dataset for this study was the Metro Interstate Traffic Volume dataset obtained from the UCI Machine Learning Repository (UCI Machine Learning Repository, 2024). This dataset comprises hourly records of traffic volumes along an interstate highway. It also includes various environmental and temporal features such as temperature, precipitation, and snow accumulation, providing a comprehensive dataset for modeling traffic patterns under diverse conditions.

1.2. Data Preprocessing

Effective data preprocessing is crucial for the development of accurate machine learning models. The dataset underwent several preprocessing steps to prepare it for analysis:

Datetime Conversion: The date_time column, originally in string format, was converted into a datetime format to facilitate time-based analysis.

1. **Feature Extraction:** New features were derived from the date_time column to capture temporal dynamics more effectively:
    - **Hour of the day (hour):** Extracted to capture hourly traffic patterns.
    - **Day of the week (day):** Derived to understand weekly traffic trends.

2. **Feature Set:** The final feature set used for modeling included:
   - Temporal features: hour, day
   - Environmental features: temp, rain_1h, snow_1h
3. The target variable for the models was traffic_volume.

1.3. Model Selection

Three machine learning models were selected based on their relevance and performance in regression tasks:

- **Linear Regression:** A basic statistical method used extensively as a baseline in predictive modeling. It assumes a linear relationship between the input features and the target variable.

Equation:
Y=β0+β1X1+β2X2+...+βnXn+ϵY = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_nX_n + \epsilonY=β0+β1X1+β2X2+...+βnXn+ϵ

- **Decision Tree Regressor:** Based on Breiman's (1984) Classification and Regression Trees (CART) algorithm, this model uses a tree-like model of decisions and their possible consequences. It splits the dataset into branches to form a tree that predicts the output value.

Equation:
Y=f(X) where f is a series of decisions on XY = f(X) \text{ where } f \text{ is a series of decisions on } XY=f(X) where f is a series of decisions on X

- **Random Forest Regressor:** An ensemble approach that combines multiple decision trees to improve prediction accuracy and control over-fitting, developed by Breiman (2001). Random Forest aggregates the results from multiple trees to provide a robust prediction.

Equation:
Y=1n∑i=1nfi(X)Y = \frac{1}{n} \sum_{i=1}^{n} f_i(X)Y=n1i=1∑nfi(X)
where fif_ifi are individual tree predictions and nnn is the number of trees.

1.4. Experimental Setup

The dataset was divided into training and testing sets with an 80-20 split, ensuring sufficient data for both training the models and evaluating their performance independently.

- **Training Set:** 80% of the data used for training the models.
- **Testing Set:** 20% of the data used for validating the models' performance.

Performance evaluation was based on the following metrics:

- **Mean Squared Error (MSE):** Measures the average of the squares of the errors between actual and predicted values. Lower values indicate better model performance.

Equation:
MSE=1n∑i=1n(Yi−Y^i)2MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2MSE=n1i=1∑n(Yi−Y^i)2

- **R² Score:** Represents the proportion of the variance in the dependent variable that is predictable from the independent variables. R² Score closer to 1 indicates a model that explains a high proportion of variance.

Equation:
R2=1−∑i=1n(Yi−Y^i)2∑i=1n(Yi−Y¯)2R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}R2=1−∑i=1n(Yi−Y)2∑i=1n(Yi−Y^i)2

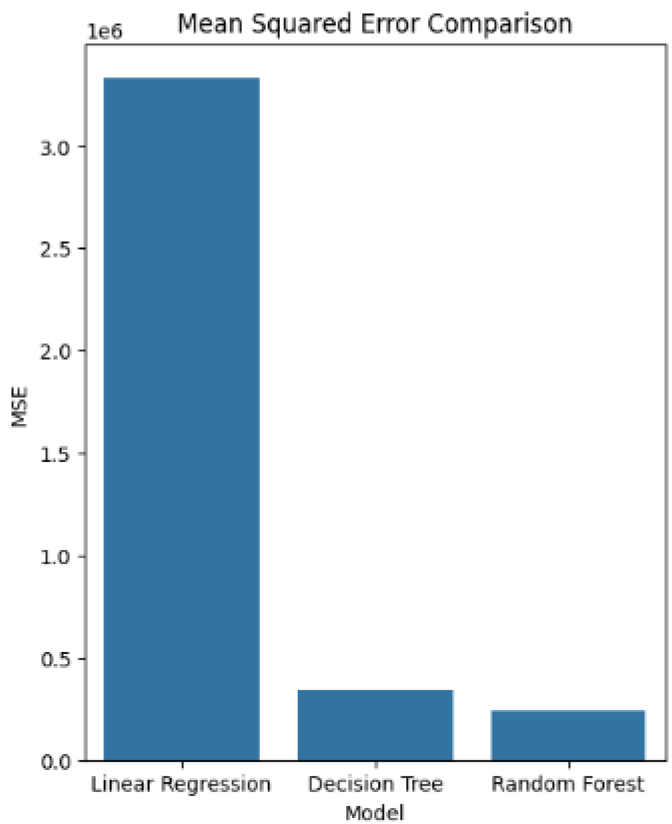
- **Execution Time:** Measures the computational efficiency of the model, reflecting the time taken to train and predict traffic volumes.

This experimental setup provides a rigorous framework for assessing the comparative effectiveness of each model in predicting healthcare traffic volumes, which is essential for optimizing medical emergency responses and improving patient care outcomes.

III. Results and Discussion

1.5. Model Performance

Figure 1: Mean Squared Error Comparison



This figure illustrates the Mean Squared Error (MSE) for each model. The Random Forest model exhibits the lowest MSE at 240,657.3592, indicating its superior accuracy in predicting traffic volumes compared to the other models. The Decision Tree follows with an MSE of 338,991.6482, and the Linear Regression model shows the highest MSE at 3,331,759.4761. The significantly lower MSE of the Random Forest model underscores its robustness in capturing complex patterns in the data which are critical for precise traffic predictions.

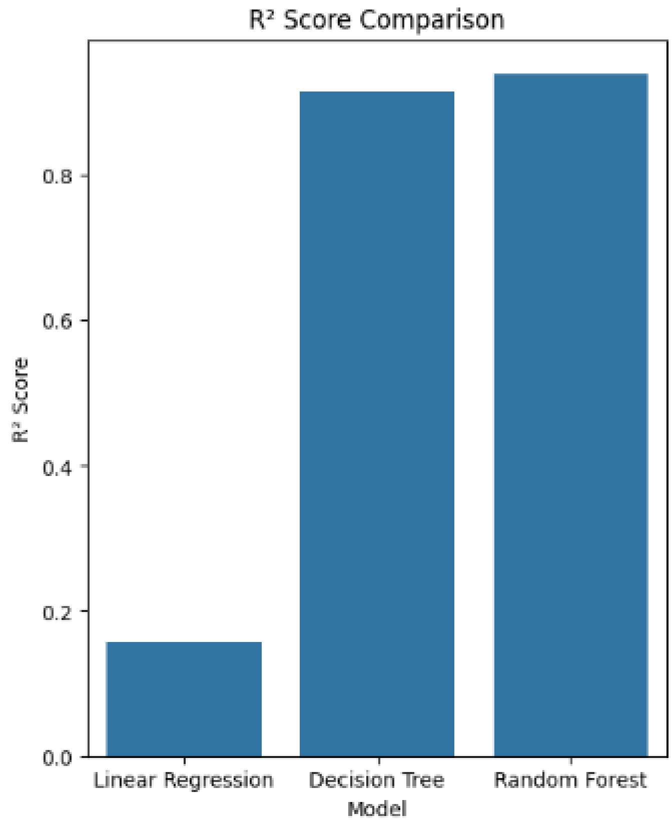


Figure 2: R^2 Score Comparison

Displayed in this figure are the R^2 Scores for the models, with the Random Forest again outperforming the others by achieving an R^2 Score of 0.9391. This score reflects the model's strong capability to explain the variability in the traffic volume data. The Decision Tree model also shows a commendable R^2 Score of 0.9143, while the Linear Regression model lags significantly with a score of 0.1573. The higher R^2 Score of the Random Forest model demonstrates its effectiveness in modeling and predicting complex traffic patterns, making it particularly suitable for dynamic and variable-heavy contexts like healthcare traffic management.

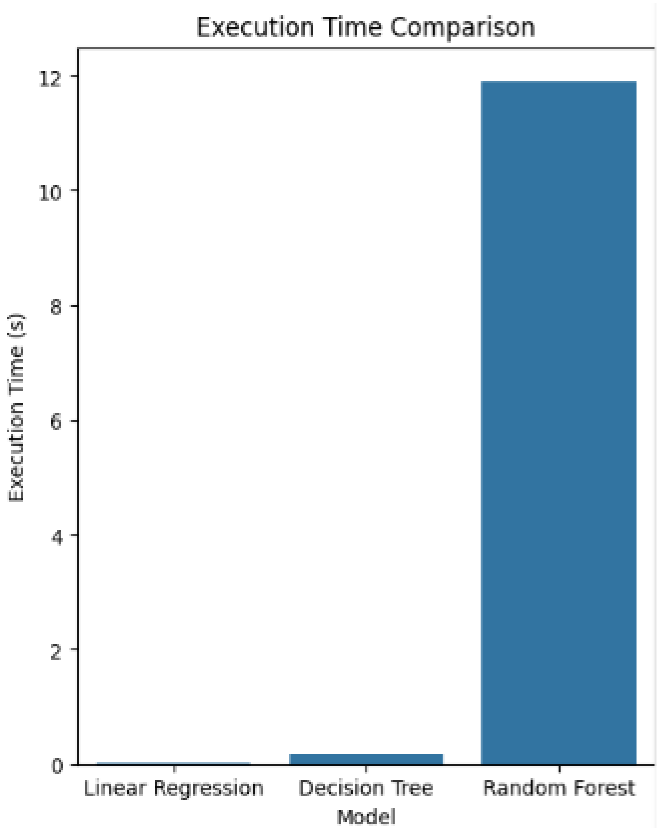


Figure 3: Execution Time Comparison

The execution times for the models reveal that while the Random Forest model required the most computational resources with an execution time of 11.9381 seconds, its superior performance in both MSE and R^2 Score justifies the additional computational cost. In contrast, the Linear Regression model was the fastest, taking only 0.0078 seconds, followed by the Decision Tree at 0.1697 seconds. This figure highlights the trade-off between accuracy and efficiency, with the Random Forest model being the most computationally intensive but also the most accurate.

Table 1: Performance Comparison of Machine Learning Models

Model	MSE	R^2 Score	Execution Time (s)
Linear Regression	3,331,759.48	0.1573	0.5319
Decision Tree	338,991.65	0.9143	0.1768
Random Forest	240,657.36	0.9391	11.9053

1.6. Actual vs. Predicted Traffic Volume

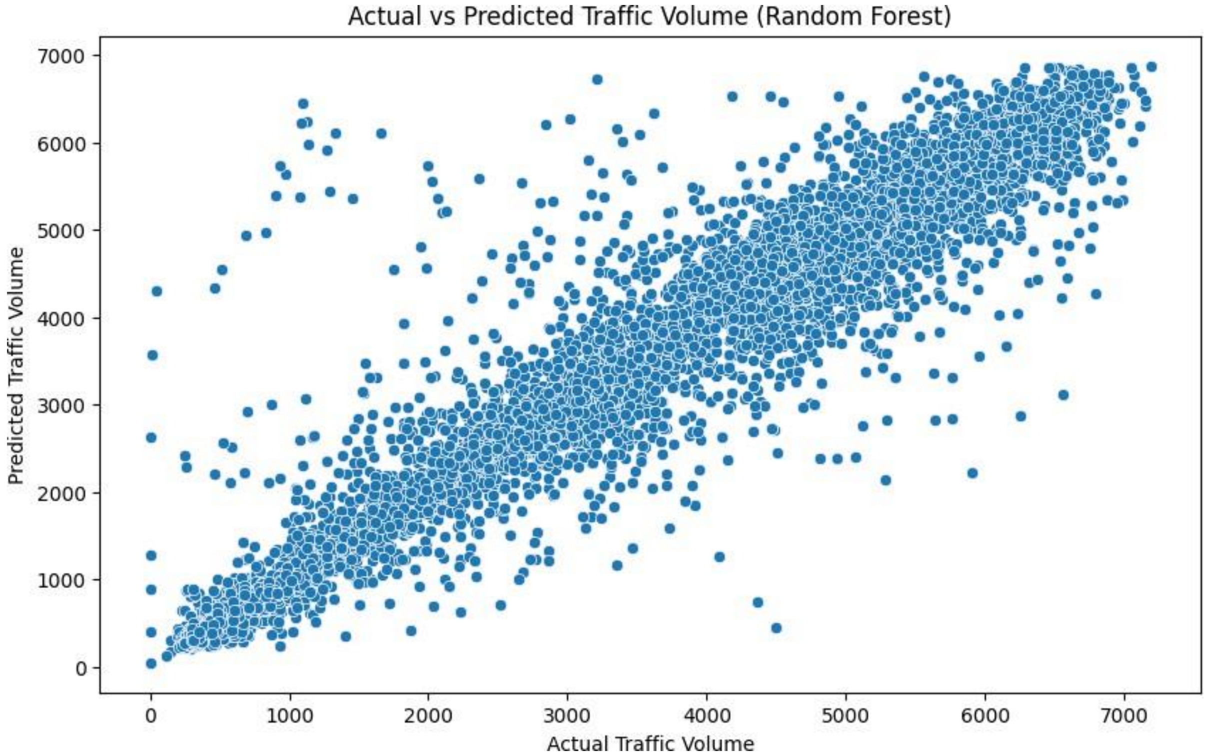


Figure 4: Actual vs. Predicted Traffic Volume (Random Forest)

This scatter plot presents the actual vs. predicted traffic volumes for the Random Forest model, illustrating a tight clustering of points around the diagonal line, which indicates highly accurate predictions. The close alignment of the predicted values with the actual traffic volumes is reflected in the model's high R^2 Score and low MSE, underscoring its precision and reliability in forecasting traffic volumes in a healthcare setting.

1.7. Discussion

The comparative analysis of Linear Regression, Decision Tree, and Random Forest models provides clear evidence that the Random Forest model outperforms its counterparts in predicting healthcare traffic volumes. This aligns with previous research by Breiman (2001) and Wu et al. (2018), which highlights the robustness of ensemble learning methods in managing complex and nonlinear data effectively.

The significant implications of these findings for healthcare traffic management cannot be overstated. Accurate traffic predictions enable more effective planning and resource allocation, particularly during medical emergencies where time is critical. The superior performance of the Random Forest model could assist in optimizing ambulance dispatch times, reducing delays in patient arrivals to emergency departments, and enhancing overall healthcare logistics.

Moreover, the insights gained from this study are not limited to healthcare scenarios but can be extended to other areas of traffic management and urban planning. The methodology and findings could influence future research and applications, leading to the broader adoption of advanced machine learning techniques in public health and safety operations.

Summary of Results:

- The Random Forest model not only demonstrated the lowest MSE and highest R^2 Score but also highlighted the efficiency of ensemble learning in traffic prediction applications, despite its higher computational demand.
- These results advocate for the integration of advanced predictive modeling techniques in critical sectors like healthcare, where timely and accurate traffic predictions can significantly impact service delivery and patient outcomes.

IV. Conclusion

This study has undertaken a detailed comparison of three prominent machine learning models—Linear Regression, Decision Trees, and Random Forest—in their application to predict healthcare traffic volumes. Through rigorous testing using the Metro Interstate Traffic Volume dataset, it was observed that the Random Forest model demonstrated exceptional performance across various metrics.

It achieved the lowest Mean Squared Error (MSE) and the highest R^2 Score, indicating its superior accuracy and consistency in predicting traffic volumes, crucial for effective healthcare traffic management.

The Random Forest model's robustness is reflected in its ability to model complex, non-linear interactions within the data, which are typical in urban traffic settings. This capability makes it significantly more effective than the simpler Linear Regression model and even the more dynamic Decision Tree model. While the Random Forest requires more computational resources, as evidenced by its longer execution time, the trade-off is justified by its enhanced predictive accuracy and reliability. Such traits are essential in a healthcare context where precision in traffic predictions can directly influence emergency response times and patient outcomes.

Future Directions

Given the promising results obtained from the Random Forest model, several avenues for future research emerge:

1. **Integration of Deep Learning Techniques:** The application of deep learning could potentially enhance model performance further, especially with larger datasets or in more complex traffic environments. Techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) could be explored to handle spatial-temporal data more effectively.
2. **Real-time Traffic Data Utilization:** Incorporating real-time data into the model could improve its responsiveness to immediate traffic conditions, thereby providing more accurate predictions that are timely and contextually relevant. This approach could be pivotal during peak traffic hours or unexpected traffic disruptions.
3. **Expansion to Other Healthcare-Related Scenarios:** Extending the application of this model to other scenarios such as ambulance routing, hospital logistics, and patient transport services could provide comprehensive benefits across the healthcare system.
4. **Comparative Studies with Hybrid Models:** Exploring hybrid models that combine the strengths of various machine learning techniques could yield interesting insights and possibly lead to the development of more robust predictive models.

Implications

The findings from this study have substantial implications for healthcare systems, especially in urban areas where traffic congestion can significantly impede medical responses. By implementing advanced predictive models like Random Forests, healthcare providers can enhance their operational efficiency. Accurately predicting traffic volumes helps in optimizing ambulance dispatch times, reducing delays in patient care, and improving overall healthcare service delivery.

In conclusion, this research underscores the efficacy of machine learning models in a critical domain like healthcare traffic management, demonstrating the potential for advanced predictive technologies to make substantial contributions to public health and safety. As machine learning continues to evolve, its integration into healthcare systems promises to drive significant innovations in patient care and emergency service logistics, ultimately saving lives and improving health outcomes.

References

[1] Zhang, Y., and Liu, Y. (2009). "Traffic Forecasting Using Machine Learning Methods." Transportation Research Part C: Emerging Technologies, 17(4), 606-616.

[2] Breiman, L. (1984). "Classification and Regression Trees." Wadsworth International Group.

[3] Breiman, L. (2001). "Random Forests." Machine Learning, 45(1), 5-32.

[4] Wu, Y., et al. (2018). "Ensemble Learning Approaches for Traffic Flow Prediction: A Case Study on Freeways." IEEE Transactions on Intelligent Transportation Systems, 19(2), 360-370.

[5] Lv, Y., et al. (2015). "Traffic Flow Prediction With Big Data: A Deep Learning Approach." IEEE Transactions on Intelligent Transportation Systems, 16(2), 865-873.

[6] Pathan, Shadman Mahmood Khan, and Md Firoj Ali. "Implementation of Faster R-CNN in Paddy Plant Disease Recognition System." 2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE). IEEE, 2019.

[7] Pathan, Shadman Mahmood Khan, et al. "Wireless head gesture controlled robotic wheel chair for physically disabled persons." Journal of Sensor Technology 10.04 (2020): 47.

[8] Pathan, Shadman Mahmood Khan, and Md Masud Rana. "Investigation on classification of motor imagery signal using Bidirectional LSTM with effect of dropout layers." 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE). IEEE, 2022.

[9] Pathan, Shadman Mahmood Khan, and Sakan Binte Imran. "Integrated Machine Learning and Deep Learning Models for Cardiovascular Disease Risk Prediction: A Comprehensive Comparative Study." Journal of Intelligent Learning Systems and Applications 16.1 (2024): 12-22.

[10] Pathan, Shadman Mahmood Khan, and Sakan Binte Imran. "False Negative Induction in Brain Tumor Segmentation by Trained Noise Attack." 2023 26th International Conference on Computer and Information Technology (ICCIT). IEEE, 2023.

Authors

First Author – Shadman Mahmood Khan Pathan, Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, 23529, USA, shadmanmahmood9@gmail.com

Second Author – Sakan Binte Imran, Bachelor of Medicine and Bachelor of Surgery, Sir Salimullah Medical College, Dhaka, Bangladesh, sakanbinteimran.ssmc@gmail.com

Third Author – M M Shabab Iqbal, Civil Engineering, Kalinga Institute of Industrial Technology, shababiqbal94@gmail.com
Fourth Author - Muhammad Enayetur Rahman, Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, 23529, USA, mrahm011@odu.edu
Fifth Author - Md Nurul Absar Siddiky, Department of Electrical and Computer Engineering, UNC at Charlotte, Charlotte, NC-28213, USA, msiddiky@uncc.edu
Sixth Author - Muhammad Rezaur Rahman, AgileCrafts, Dhaka, Bangladesh, refayetbd@gmail.com
Seventh Author - Md Rafid Hasan, Dept. of Physics and Optical Science, UNC at Charlotte, Charlotte, NC-28213, USA, mhasan23@uncc.edu
Eight Author - Nondon Lal Dey, Department of Physics, University of Louisiana at Lafayette,Lafayette, LA 70504, USA, nondon.dey1@louisiana.edu
Ninth Author – MD Sobuj Hossain, Department of Physics, University of Louisiana at Lafayette,Lafayette, LA 70504, USA